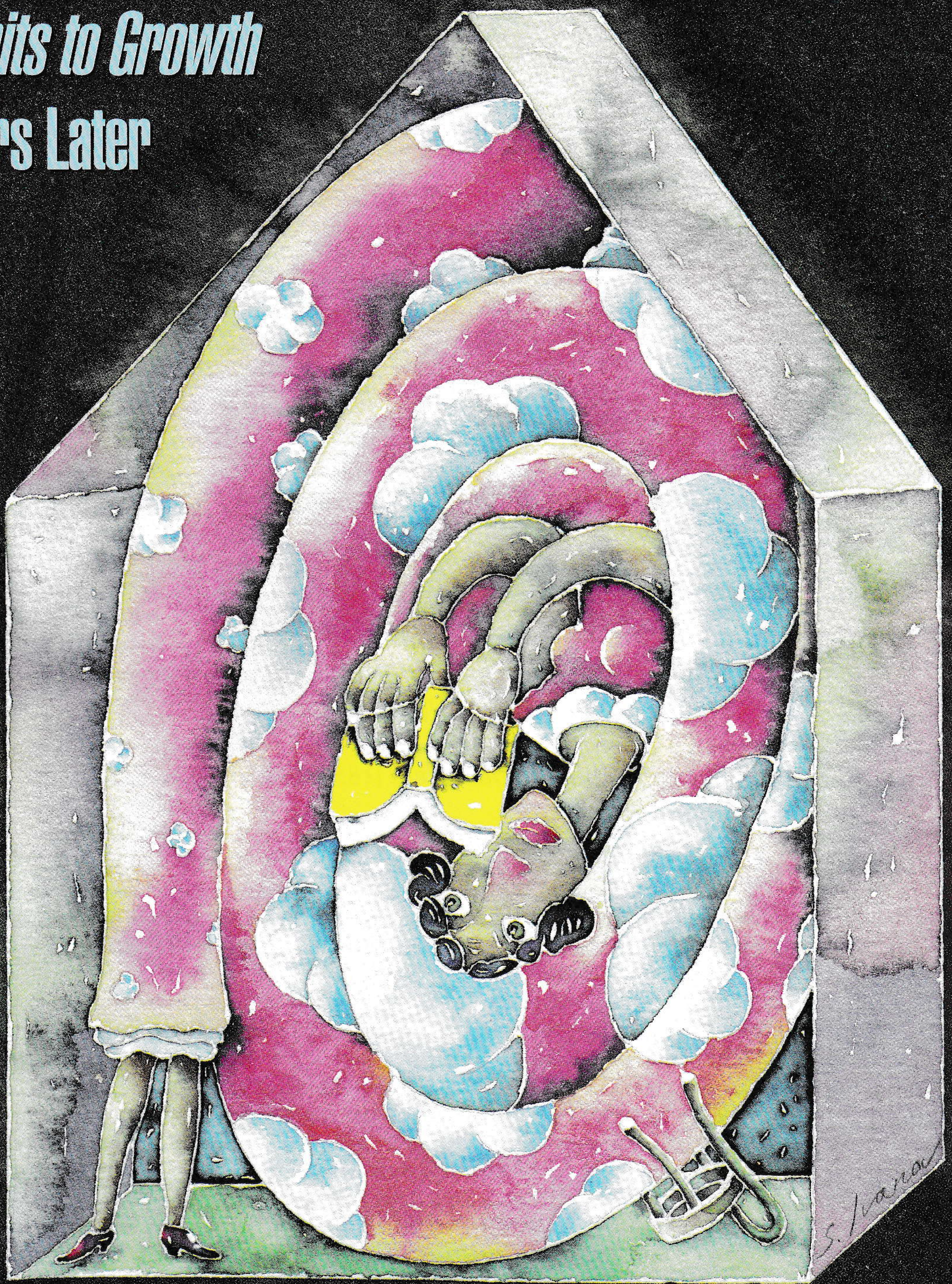
QUANTUM

SEPTEMBER/OCTOBER 1997

US $7.50/CAN $10.50

The Limits to Growth
25 Years Later



Springer

GALLERY Q

Catlin and Indian Attacking Buffalo Herd (1857/1869) by George Catlin

GEORGE CATLIN (1796–1872) BELONGED TO THAT legion of creative spirits who left a profession (in his case, the law) to take up the pen or brush (in his case, both). After a brief stint as a portrait painter, Catlin headed west to follow a boyhood interest: Indians. Intent on documenting what he perceived as a vanishing way of life, he produced more than 500 paintings and sketches, and in 1841 he published his two-volume *Letters and Notes on the Manners, Customs, and Condition of the North American Indians.*

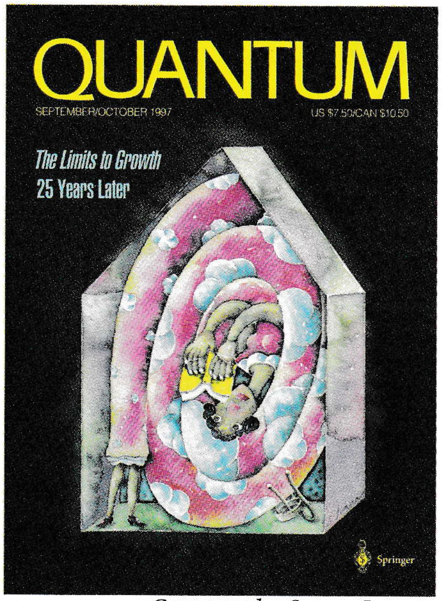
The picture presented here offers an ironic commentary (unintentional, no doubt) on the threatened lifestyle Catlin was recording—the Indian with his bow and arrow, Catlin with his rifle. When the Europeans set foot on the North American continent, an estimated 60,000,000 bison (or buffalo) roamed the western plains. The Indians of that region depended on the bison for their livelihood. By 1900, the bison had been hunted to the verge of extinction.

William F. ("Buffalo Bill") Cody won renown as a killer of bison. In the 1860s, he had been hired to provide meat for the workers on the Union Pacific Railroad, and in an eight-month period he shot 4,280 buffalo. He was not alone in his hunting. Others killed the animal for its hide, for sport, or sometimes just for its tongue. Through concerted action by cattle ranchers and conservationists, the bison was brought back from the brink and now thrives on managed government preserves (in much smaller numbers).

The story of the American bison is an instructive case of population growth, predator–prey relationships, and resource management. It is fitting that it be told (albeit briefly) in this special issue of *Quantum*.

QUANTUM

SEPTEMBER/OCTOBER 1997 VOLUME 8, NUMBER 1



Cover art by Sergey Ivanov

The woman on our cover believes in expanding her mind with a good book, but somehow her body took a metaphorical turn of its own. What book is she reading? Why, *The Limits to Growth*, the international bestseller of 1972.

This issue of *Quantum* is devoted to the legacy of that pathfinding report to the Club of Rome. Our cover is but the prelude to a suite of allegorical paintings that accompany each of the five feature articles.

Turn to the Front Matter on page 2 for a brief introduction to our 25th Anniversary commemoration of *The Limits to Growth*.

FEATURES

DEPARTMENTS

Indexed in *Magazine Article Summaries, Academic Abstracts, Academic Search, Vocational Search, MasterFILE*, and *General Science Source*. Available in microform, electronic, or paper format from University Microfilms International.

The unlimited appeal of *The Limits to Growth*

"Although there are 'limits' to a certain type of activity, there are no limits to learning and creativity."
—The Club of Rome

THIS ISSUE OF *QUANTUM* IS largely devoted to marking an event that occurred a quarter of a century ago—before many of our readers were even born. What was so important about the publication of *The Limits to Growth* in 1972?

To begin to answer this question, we need to back up another five years. In 1967, an Italian industrialist by the name of Aurelio Peccei and a Scottish scientist, Alexander King, saw that governments worldwide seemed incapable of addressing certain long-range trends that threatened the well-being of future generations. They saw the world's population increasing at a staggering pace, nonrenewable resources being depleted at increasing rates, and millions of people living at subsistence levels or under threat of famine with little prospect of improvement. The time had come, they felt, to mobilize like-minded people and take action to build a saner and more sustainable world.

In April of the next year, thirty-six European economists, scientists, and statesmen met in Rome to discuss these issues unconstrained by politics or ideology. They continued to meet in other cities and expanded their ranks, but the location of their first meeting had given them a name: The Club of Rome.

In 1969, the members decided that a quantitative model of the present and future "predicament of mankind" might prove more persuasive with decision makers than their verbal statements had been. In 1970 the Club of Rome turned to Jay Forrester of MIT, who had developed a computer technique for dealing with the complexities of industrial production and had already applied it to social questions as well. In late 1970 funding was secured, and an international staff of seventeen specialists at MIT, under the direction of Dennis Meadows, began work on a world model based on Forrester's concept of *system dynamics*. In addition to the technical reports presented at meetings in 1971, the MIT team produced a book intended for the general public, *The Limits to Growth*, written by Donella Meadows.

The Limits to Growth was immediately hailed by many as revolutionary, teaching us how to see the Earth as a closed system and, with its sophisticated interconnections, equations, data tables, giving us the courage to address problems that had seemed impossibly complex. It was also reviled by others who pointed out the imprecision of the variables, or the ingenuity of human beings in overcoming limits, or the apparent failure of previous models (most notably the one offered by Thomas Malthus in 1798). Ironically, the book was born into a world riding a 20-year crest of economic growth. Who needed its (supposed) "gloom and doom"? Two years later, however, an oil embargo sent a shudder through the industrial world and cast the question of energy production and consumption in a whole new light.

The debate sparked by *The Limits to Growth* continues, as researchers revise their models and update their data in support of "progrowth" or "sustainable" economies. Many are now focusing more on critical assessments of economic models and less on Forrester-type "world models." The Club of Rome itself has changed its emphasis, choosing to focus on discrete aspects of global problems (for example, international law regarding the use of the seas). Its most recent monograph, *Taking Nature into Account*, criticizes the kind of "national accounting" that makes the gross national product (GNP) such an influential (and distorted) measure of economic success.

To return to our initial question: why is the publication of *The Limits to Growth* worth commemorating? It marked the first time that a global model of this sort had been commissioned by an independent body (rather than a government or the United Nations). But perhaps more importantly, it was the first study to make an explicit link between economic growth and consequences for the environment. It questioned the reigning dogma that "growth is good" and forcefully inserted the concept of *sustainability* into our policy debates.

We hope this issue of *Quantum* will give you some sense of the excitement and hope arising from work with system dynamics and world modeling. Our sampling of this vast subject area consists of five feature articles:

- "*The Limits to Growth* Revisited" offers a primer on exponential growth, overshoot, and dynamic modeling.
- "Overshooting the Limits" demonstrates different ways a limit can be approached and the consequences of overshooting it.
- "The World in a Bubble" takes us inside the simulated Earth of Biosphere 2, where sustainability became a life-or-death issue.
- "Learning from a Virus" applies the techniques of system dynamics to the spread of an illness in a population.
- "Input–output Economics" describes an effort to harness I–O economics to take account of the "bads" produced by an economy as well as the nonrenewable resources required to sustain it.

Our Kaleidoscope in this issue is a graphical representation of the revised World3 model that appears in *Beyond the Limits*, the 1991 sequel to *The Limits to Growth*. The model, along with its supporting equations and data, is included in the freely distributed modeling software Vensim PLE (see page 19).

The *Quantum* staff would like to thank Kurt Kreith of the University of California–Davis for his unflagging efforts as the editor of this special issue. We welcome comments from our readers. Our e-mail address is quantum@nsta.org; our postal address is *Quantum*, 1840 Wilson Boulevard, Arlington VA 22201-3000.

—Tim Weber,
Managing Editor

QUANTUM

THE MAGAZINE OF MATH AND SCIENCE

A publication of the National Science Teachers Association (NSTA)
& Quantum Bureau of the Russian Academy of Sciences
in conjunction with
the American Association of Physics Teachers (AAPT)
& the National Council of Teachers of Mathematics (NCTM)

The mission of the National Science Teachers Association is to promote excellence and innovation in science teaching and learning for all.

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 8 (6 issues) will be published in 1997–1998. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau (Moscow, Russia), as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington VA 22201-3000; telephone: (703) 243-7100; electronic mail: quantum@nsta.org. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York NY 10010-7858.

Printed on acid-free paper.

Periodicals postage paid at New York, NY, and additional mailing offices. **Postmaster:** send address changes to: *Quantum*, Springer-Verlag New York, Inc., Journal Fulfillment Services Department, P. O. Box 2485, Secaucus NJ 07096-2485.

Subscription Information:

North America: Student rate: $17.70; Personal rate (non-student): $25; Institutional rate: $45; Single Issue Price: $7.50. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Send orders to *Quantum*, Springer-Verlag New York, Inc., P.O. Box 2485, Secaucus NJ 07096-2485; or call 1-800-SPRINGER (777-4643) (in New York, call [201] 348-4033).

All Countries Outside North America: Personal rate is US$37 including surface delivery charges; institutional rate is US$57; airmail delivery is US$18 additional (all rates calculated in DM at the exchange rate current at the time of purchase). SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Orders may be placed through your bookseller or directly through Springer-Verlag, Postfach 31 13 40, D-10643 Berlin, Germany.


the sooner they begin, the greater will be their chances of

The Limits to Growth revisited

A bit of history, and a challenge to our readers

by Kurt Kreith

IT HAS BEEN 25 YEARS SINCE The Club of Rome published its provocative study *The Limits to Growth.* To commemorate this event the editors of *Quantum* have undertaken a special issue, one that describes *The Limits to Growth* and examines its implications. It also invites *Quantum* readers to engage in some "dynamic modeling" as described in this and subsequent articles.

Contrary to the usual *Quantum* fare, my introduction to *The Limits to Growth* (let's call it LTG for short) will not deal with an established scientific phenomenon or mathematical structure. Rather, it presents LTG's framework for thinking about the Earth's ability to sustain both humankind and our industrial economy. Also we will become familiar with some remarkable software that can help us develop our own framework for thinking about what The Club of Rome's founder referred to as "the predicament of mankind."

But before turning to LTG itself, it may be useful to recall some other historical events. When Nicolaus Copernicus published his monumental *De Revolutionibus Orbium Celestium* ("On the Revolution of the Celestial Spheres"), he also confronted the world with a controversial new framework for thinking about a system, namely the solar system comprised of the earth, sun and its planets. The immediate question, "Is Copernicus correct?" was, in retrospect, not the most important question to ask. The Copernican model, one in which the Earth and planets pursue circular orbits about a stationary Sun, is in detail not correct. Nonetheless, Copernicus's heliocentric framework for thinking about celestial change was a monumental step, one that set the stage for Kepler, Newton, and others to apply the finishing touches.

Art by Sergey Ivanov

Similarly, when Charles Darwin published *Origin of Species,* he sought to explain the diversity of life forms on earth without reference to modern principles of heredity (to say nothing of DNA). Nonetheless, Darwin's theory of natural selection provided an important new framework that enabled others to think more creatively about the natural world. Without claiming that LTG is destined to play a comparable role in understanding the environmental issues confronting our civilization, it is worth noting that science often advances in a less than orderly fashion.

Such historical examples also suggest that a 25th anniversary may be somewhat premature for efforts to resolve the controversies that surround LTG. It was more than a hundred years before Newton provided the Copernican model with its final vindication, showing that the motion of the planets follows from an inverse square law of gravitation. And while Darwinism continues to command great respect, the precise role of natural selection in evolutionary change remains a topic of controversy to this day.

Accordingly, this issue of *Quantum* is at best an "interim report" on *The Limits to Growth.* Aside from describing the framework that LTG provides for thinking about "the predicament of mankind," it is also intended to provide *Quantum* readers with an introduction to the field of "system dynamics." Indeed, it is LTG's compelling arguments for the need to address issues of environmental change in the context of "the global system" that may turn out to be LTG's most lasting legacy.

The global system

So what is the global system that LTG so forcefully called to our attention? In a nutshell, it calls for an embedding of human activity within the earth's ecosystem and then taking account of the many connections and interactions that exist. The overall idea is expressed

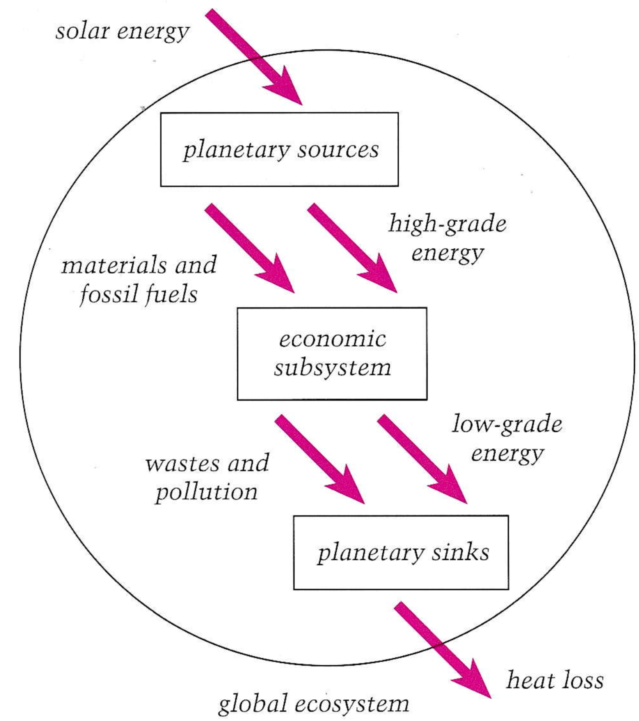


Figure 1

in figure 1 (reproduced from a 1992 sequel to LTG called *Beyond the Limits*).

Such a diagram raises topics from the physical sciences that have been addressed in past *Quantum* articles. The large circle can be thought of as representing the *boundary* of a physical system, one that is open to energy flows but (neglecting an occasional meteorite or water molecule) closed to material flows. Such a *closed system* is subject to the laws of thermodynamics and can therefore be studied in terms of established physical principles. In the July/August. 1996 issue of *Quantum* ("The Power of the Sun and You"), V. and T. Lange point out that, at a distance of 93 million miles, a square meter oriented perpendicular to the sun's rays receives 1.4 kJ of solar radiation each second. In the November/December, 1995 issue ("Less Heat and More Light"), Y. Amstislavsky reviews the physical laws governing thermal radiation (or "heat loss" in the above diagram). And in the March/April, 1991 issue ("Atmospherics"), A. Byalko sets forth the physical laws governing the Earth's thermal equilibrium, a subject of growing interest in connection with global climate change.

While such a thermodynamic analysis of the global ecosystem may be lurking in the background, it is the middle rectangle labeled "Economic Subsystem" that is at the heart of LTG. For here is a *subsystem* of the closed global ecosystem, one that is subject to human decision making. While physical systems tend to be *deterministic* in nature, this economic subsystem is believed to be subject to *free will.* It is, after all, the decisions we make as individuals and through our governing institutions that are at the heart of our efforts to address environmental issues.

Focusing on this economic subsystem, LTG's developers singled out three major components: *human population*, *food*, and *industrial production*. Going on to identify "planetary sources" with *resources* and "planetary sinks" with *pollution*, one arrives at the five variables whose interaction was studied in LTG.

In mathematical terms it is tempting to represent these five LTG variables as vertices of a regular pentagon (fig. 2). A daunting task facing LTG's authors was the formulation of "functional relationships" among the variables in figure 2, ones that specify how changes in one variable will affect the others. Some of these relationships are familiar ones, such as the dependence of population on food and the dependence of industrial production on resources. However others, such as the dependence of food on pollution and on industrial production, tend to be more subtle. Formulating such functional relations and then translating them into language understandable by a computer is the essence of the "World3" model underlying LTG.

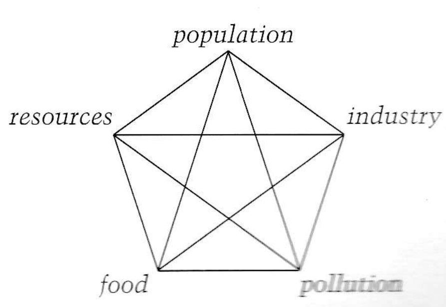


Figure 2

Why was LTG controversial?

In 1972 the MIT Project Team headed by Dennis Meadows published a nontechnical report entitled *The Limits to Growth.* After describing the exponential growth that has marked human population and industrial activity throughout the 20th century, LTG presented a series of "scenarios" generated by the World3 model. It also presented three conclusions, the first of which can be broken into two parts:

1a. If the present trends in world population, industrialization, pollution, food production and resource depletion continue unchanged, the limits to growth on this planet will be reached sometime within the next one hundred years.

1b. The most probable result will be a rather sudden and uncontrollable decline in both population and industrial capacity.

My reason for breaking conclusion 1 into two parts is to separate the controversial from the mundane. Simple "back of the envelope" calculations give credence to conclusion 1a. At current exponential rates of growth (a doubling every 40 years), the world's human population would exceed 30 billion in 2100. This corresponds to about 1/10 acre of arable land per person on earth—compared to the almost 2 or so acres presently used to sustain the diet and lifestyle of the average American. If there are individuals who wish to challenge conclusion 1a, we need only change "one hundred years" to "two hundred years." But rather than dwell on this point, let us accept LTG's assertion that exponential growth cannot be indefinitely sustained

But what about conclusion 1b? Is it unduly alarmist? Might not the approach of limits simply lead to a *reduction* in growth, one that corresponds to a smooth leveling off of human population and industrial production? On what basis did LTG raise the specter of *"a rather sudden and uncontrollable decline in both population and industrial capacity"* as a likely scenario?

Critics raised other questions as well. To what extent do the functional relationships connecting LTG's five variables reflect the complexity of "the real world"? In what sense did the insights developed in working with World3 provide a basis for LTG's conclusions? It was here that a range of important questions about the nature and significance of mathematical modeling came to the fore.

Such questions were not unanticipated. LTG's World3 model was preceded by an earlier "world model" described in Jay Forrester's book *World Dynamics.* Both LTG and *World Dynamics* stressed the distinction between *predictions* and *scenarios,* noting that computer models tend to generate the latter. Also, these authors did not claim to represent "the real world" in terms of just five variables. They pointed out, however, that all human decisions are based on some sort of mental model and that, unlike World3, many of the assumptions that go into our mental models are not explicitly spelled out. By taking into account phenomena that are neither intuitive nor part of our experience, LTG's authors believed that World3 did provide new insights, ones that can serve as a valuable adjunct to our mental models.

Quantum readers interested in pursuing this debate may want to begin by reading *The Limits to Growth* or its 1992 sequel *Beyond the Limits.* They will then be in a position to confront the arguments of LTG's critics (both sequels and critiques are listed at the end of this article). For our purposes, however, it will be more productive to sidestep this debate and focus on the underlying ideas from *system dynamics.* For once we are familiar with some of the modeling techniques arising in LTG, we will be in a better positions to reach our own conclusions about the uses and limitations of World3 and of mathematical models in general.

But before setting LTG aside, let's note its two other conclusions. Conclusion 2 reflects the fact that World3 is not inherently apocalyptic. That is, by changing some of the "inputs" that underlie LTG's more pessimistic scenarios, World3 is capable of generating scenarios that embody "sustainability." In this vein, LTG's second conclusion states:

2. It is possible to alter these growth trends and to establish a condition of ecological and economic stability that is sustainable far into the future. The state of global equilibrium could be designed so that the basic needs of each person are satisfied and each person has an equal opportunity to realize his individual human potential.

The third and final conclusion is one to which we shall return after doing some modeling ourselves. It asserts:

3. If the world's people decide to strive for this second outcome rather than the first, the sooner they begin working to attain it, the greater will be their chances of success.

Enter STELLA

In the course of their development of World3, the authors of LTG made use of ideas from system dynamics that are now embodied in several kinds of computer software. Among these is a remarkable icon-based simulation software package called STELLA®. By way of setting the stage for some of the other modeling projects in this issue of *Quantum,* it will be useful to describe some of the features of STELLA.

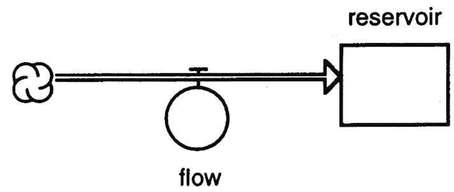


Figure 3

STELLA modeling is based on four icons, which enable us to represent dynamical systems in an intuitive format. By way of illustration, we begin with just two icons. One is a rectangle, representing a *reservoir* (or *stock*), and the other is a pipe with an arrow at one end called a *flow* (fig. 3).

The simple system represented in figure 3 can be interpreted as a tank being filled with water. If we measure volume in terms of *gallons* and time in terms of *minutes,* the flow would be measured in terms of *gallons/minute.*

One of STELLA's most important capabilities is the following: *given a specified flow (in terms of gallons per minute), calculate the amount (gallons) of water accumulated in the tank in T minutes.*

If the flow is constant (say, 3 gallons/minute), the answer is easy. The amount of water accumulated in the tank is given by $3T$, where 3 is the rate of flow and T is the amount of time elapsed. If, however, the flow varies with time, the problem becomes more challenging. Denoting a nonconstant flow by $f(t)$, we might now approximate the amount of water accumulated at time T by (1) monitoring the rate of flow at given time intervals, (2) assuming the flow is essentially constant between such monitoring, and (3) calculating the resulting accumulation by applying the rule "volume = rate × time" in these small time intervals.

Let's illustrate this idea in the case of a tank that is being filled at $f(t) = 2t + 1$ gallons/minute for 4 minutes. $A(t)$ will denote our approximation for the amount of water accumulated in the tank (fig. 4).

Time	$f(t)$	$A(t)$
$0 \le t \le 1$	1	t
$1 \le t \le 2$	3	$1 + 3(t-1)$
$2 \le t \le 3$	5	$4 + 5(t-2)$
$3 \le t \le 4$	7	$9 + 7(t-3)$

Figure 4

If we are only interested in calculating $A(t)$ at the integer times $t = 1$, 2, 3, and 4, we could do the same calculation by means of a simple spread sheet program (see "Look, Ma—No Calculus!" in the November/December 1994 issue of *Quantum*). This spreadsheet corresponds to repeated applications of the rule

$$A(T) = A(T-1) + f(T-1) \qquad (1)$$

for $T = 1, 2, 3$, or 4 (fig. 5).

Problem 1. Verify that equation (1) corresponds to $A(T) = A(0) + f(0) \cdot 1 + f(1) \cdot 1 + \ldots + f(T-1) \cdot 1$, where $A(0)$ is the volume of water in the tank at $t = 0$.

Problem 2. Apply the scheme in figure 4 (or figure 5) to the flow $f(t) = 1 + t^2$ to calculate $A(4)$.

In case greater accuracy is desired, we could monitor the flow into the tank more frequently than once a minute. Monitoring such flows four times a minute calls for repeated applications of the rule $A(t) = A(t-0.25) + f(t-0.25)/4$. As in problem 1 above, this leads to

$$A(T) = A(0) + \frac{f(0)}{4} + \frac{f(0.25)}{4} + \frac{f(0.5)}{4} + \frac{f(0.75)}{4} + \cdots + \frac{f(T-0.25)}{4}.$$

Those familiar with calculus may notice that $A(T)$ is just a "Riemann sum corresponding to the integral of $f(t)$ from $t = 0$ to $t = T$." In other words, lurking underneath STELLA's friendly icons is a spreadsheet program that performs numerical integration and thereby provides approximate solutions to problems such as

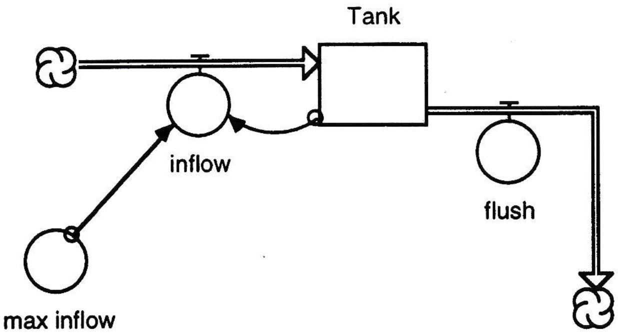


Figure 6

$$\frac{dA}{dt} = f(t);$$
$$A(0) = A_0.$$

However, one of the beauties of such software packages is that they enable us to conceptualize and "run" dynamical systems without involving the calculus. This capability also enables us to do some rather sophisticated modeling in an icon-based format, one that requires only algebra and functional notation.

By way of introducing STELLA's two other icons, let's consider a mundane example: the tank that sits on the back of the ubiquitous flush toilet! This tank has both an outflow (for creating a siphon effect in the toilet) and an inflow (for refilling the tank). Inside the tank there is a "float," one that links the rate of inflow to the amount of water in the tank and is responsible for shutting off the inflow once the tank has acquired a prescribed level. In modeling such a system, STELLA uses a *connector* to reflect the fact that **inflow** depends on the amount of water in the **Tank** (fig. 6).

Also included in figure 6 is a circular icon called a *converter* that can be used to introduce a variable or "parameter" that affects the system's dynamics. In the context of figure 5, the converter called **max inflow** is connected to **inflow**. What we have in mind is the fact that toilet tanks have shutoff valves whose setting controls the maximum rate at which the tank can be refilled.

Having created such a dynamical system, STELLA's dialog boxes enable us to program rules such as the following into a spreadsheet underlying these icons:

1. At time $t = 0$, the **Tank** contains 11 gallons of water—that is, $A(0) = 11$.
2. At time $t = 1$, a flush drains the tank at 7 gallons/s for 2 seconds.
3. The **max inflow** converter is set at 2 gallons/s.
4. The **inflow** is given by the smaller of **max inflow** and **11 – Tank.**

When this symbolic toilet is flushed, the water in the tank falls rapidly for two seconds. As soon as the volume falls below 11 gallons, the tank begins refilling, as defined by **inflow** in rule 4, until it again reaches its capacity of 11 gallons.

Not only does STELLA calculate the changes in the variables highlighted above, it will also draw a graph describing the dynamics of such a system (fig. 7).

As you may have guessed by now, these computational techniques can be applied to phenomena other than just tanks and water flows. A reservoir can represent a population, with flows in and out representing births and deaths. A reservoir can also represent money in the bank, with flows in and out representing deposits and withdrawals. Indeed, a host of different variables that are subject to increase and decrease according

	A	B	C
1	Time	f(t)	A(t)
2	0	=2*A2+1	0
3	=A2+1	=2*A3+1	=C2+B2*1
4	=A3+1	=2*A4+1	=C3+B3*1
5	=A4+1	=2*A5+1	=C4+B4*1
6	=A5+1	=2*A6+1	=C5+B5*1

Figure 5

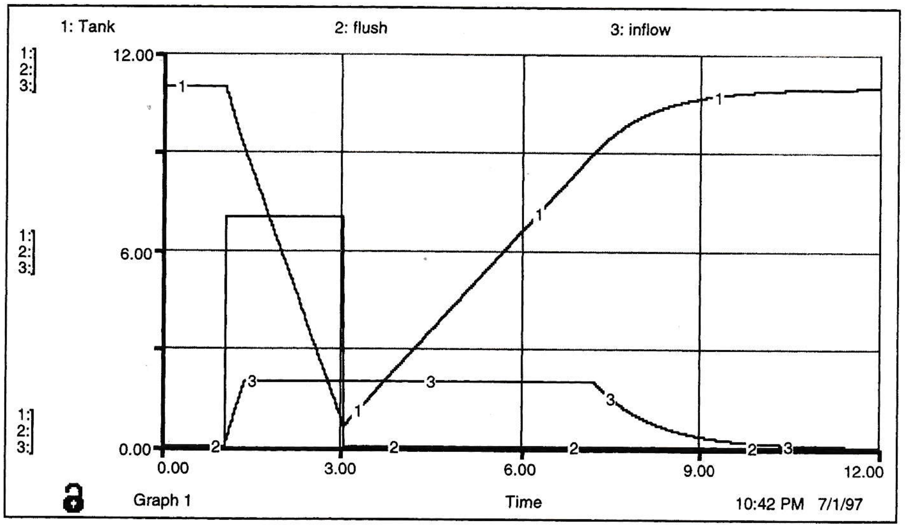


Figure 7

to some mathematical rule can be represented by such an icon. Furthermore, the rate at which a variable changes can depend on that variable itself (this is called *feedback)* and on the state of other variables (this occurs in a *system*).

Against this background, we can now think of elaborating on figure 2 in STELLA format. The vertices of the pentagon would be replaced by five "sectors" containing reservoirs with numerous inflows and outflows. The ten edges would be replaced by a bevy of connectors and converters that reflect functional relationships between population, food, pollution, industrial capacity, and resources. The resulting iconic diagram would provide a detailed blueprint for anyone seeking an explicit understanding of the assumptions that underlie this particular "world model."

When LTG appeared in 1972, very few people were in a position to understand such models, their uses, and their limitations. By contrast (as we'll see below), software such as STELLA now makes the underlying ideas accessible to anyone with a modern desktop computer.

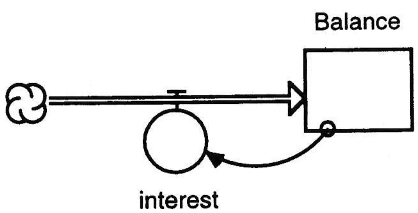


Figure 8

Population dynamics

In "Look, Ma—No Calculus!" (November/December 1994), spreadsheets and banking analogies were used to describe some important ideas from population dynamics. Here "$100 deposited in a bank at 10% interest per year" was the starting point for developing the concept of exponential growth, while variations on this banking theme led to other phenomena. Let's now consider how STELLA enables us to deal with these same ideas and, in particular, with investigation 2 as posed in "Look, Ma—No Calculus!"

A balance of $100 deposited at 10% interest corresponds to the STELLA icons shown in figure 8. Here we must also use STELLA's dialog boxes to specify that the initial **Balance** is 100 and that the **interest** is **0.1** · **Balance**. Given simple interest paid annually, this leads to the results shown in figure 9.

However, STELLA can also be asked to compound your interest quarterly, implementing the rule $B(t + 0.25) = B(t) + 0.025 \cdot B(t)$ (fig. 10). This is done by setting $DT = 0.25$ in STELLA's "Time Specs" menu. (On most older versions of STELLA $DT = 0.25$ is the default setting. To get annual compounding one must set $DT = 1$.)

Continuing in this vein, investigation 2 from "Look, Ma—No Calculus!" was based on a dubious Murky Savings and Loan, one that offers its clients an attractive 10% rate of interest. However, Murky also subjects their accounts to a "very small" service fee of 0.05%, albeit one that is applied to the *square* of the balance! In terms of STELLA's icons, this means that the reservoir **Balance** also has an outflow attached to it, which we'll call **fee** (fig. 11).

The connectors from **Balance** to both **interest** and **fee** enable us to specify that **interest = 0.1 · Balance**, while **fee = 0.0005 · Balance · Balance**. With dialog boxes so defined, STELLA readily provides a graph describing the growth of $100 deposited at Murky (fig. 12).

To see why an initial deposit of $100 never grows past $200 at Murky, let $B(t)$ denote the balance after t

10:51 PM 7/1/97

Time	Balance	
0	$100.00	
1	$110.00	
2	$121.00	
Final	$133.10	

Figure 9

10:55 PM 7/1/97 Table 1

Years	Balance		
.00	$100.00		
.25	$102.50		
.50	$105.06		
.75	$107.69		
1.00	$110.38		
1.25	$113.14		
1.50	$115.97		
1.75	$118.87		
2.00	$121.84		
2.25	$124.89		
2.50	$128.01		
2.75	$131.21		
Final	$134.49		

Figure 10

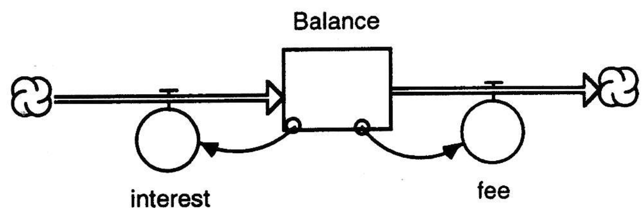


Figure 11

years. Assuming annual compounding, Murky's rule can be written

$$B(t+1) - B(t) = 0.1 \cdot B(t) - 0.0005 \cdot B(t) \cdot B(t)$$
$$= 0.0005 \cdot (200 - B(t)) \cdot B(t).$$

This formulation shows that balances under $200 will grow in value, but that balances over $200 will actually decline in value. At $B = 200$ the annual service fee of $0.0005 \cdot B^2$ exactly cancels out the annual interest of $0.1 \cdot B$.

Problem 3. Given quarterly compounding, $B(t + 1/4) - B(t) = 0.025 \cdot B(t) - 0.000125 \cdot B(t) \cdot B(t)$. What will happen to $50 deposited at Murky under this rule?

While unlikely in the world of banking, these ideas do occupy an honored place in population dynamics. During the 19th century a Belgian biologist named Verhulst suggested that in certain constraining situations (for instance, yeast cells in a closed jar) populations do grow according to the rule $dN/dt = a \cdot N(t) - b \cdot N(t)^2$, where the constant a reflects the population's unfettered growth rate and b reflects the severity of some external constraint to continued growth. Given annual compounding, dN/dt corresponds to $N(t + 1) - N(t)$, and Verhulst's equation becomes

$$N(t+1) - N(t) = a \cdot N(t) - b \cdot N(t)^2$$
$$= b \cdot \left[\frac{a}{b} - N(t)\right] \cdot N(t).$$

The fact that such a population increases if $N < a/b$ and decreases if $N > a/b$ suggests that a/b can be interpreted as the "carrying capacity" of the system being represented. That is, if $N(0) < a/b$, such a population will grow, but never past its limiting value of a/b. In LTG such S-shaped functions are said to correspond to *sigmoid growth.*

Of course we should not take such models too seriously. The rule $N(t + 1) - N(t) = a \cdot N(t) - b \cdot N(t)^3$ also leads to sigmoid growth—but toward a different limit. To see this, you need only calculate what happens if Murky applied its 0.05% service fee to the *cube* of your balance! What is important here is that a linear feedback "$a \cdot N(t)$" is being offset by an opposite effect "$-b \cdot f(N(t))$," where $f(N)$ is of the form N^p, $p > 1$. Such "superlinear damping" terms also give rise to a form of sigmoid growth whose limits are determined by a and b.

Problem 4. Find the "carrying capacity" of an environment in which a population grows according to $N(t + 1) - N(t) = 0.1 \cdot N - 0.0005 \cdot N^3$. Repeat for $N(t + 1) - N(t) = 0.1 \cdot N - 0.0005 \cdot N^{3/2}$.

Back to LTG

While LTG's World3 addressed the interaction of five different variables, the Verhulst equation deals with only one. Yet even such a one-dimensional "shadow" of World3 can provide important insight into the phenomena arising in LTG. Suppose that after finding that figure 9 generates exponential growth (of the kind that has marked human population and industrial production throughout the 20th century), we have done some back-of-the-envelope calculations indicating that such exponential growth cannot be sustained throughout the 21st century. Then, going back to Verhulst's time-honored ideas from population dynamics, we have also hypothesized a form of "superlinear damping" of the form $-bN^2$, perhaps reflecting the effects of pollution and declining resources on the global ecosystem.

Well, so far so good. The sigmoid growth given by Verhulst's model corresponds to a "soft landing," whereby a population stabilizes at its carrying capacity $N = a/b$. (The important fact that carrying capacity may itself change with time is addressed in equations (2) and (3) below.) This raises the following important question: *What was it about World3 that corresponded to "a rather sudden and uncontrollable decline in both population and industrial capacity"?*

Investigation 2 from "Look, Ma—No Calculus!" suggests that at least part of the answer lies in a phenomenon called "delayed feedback." Recall that in programming the dialog

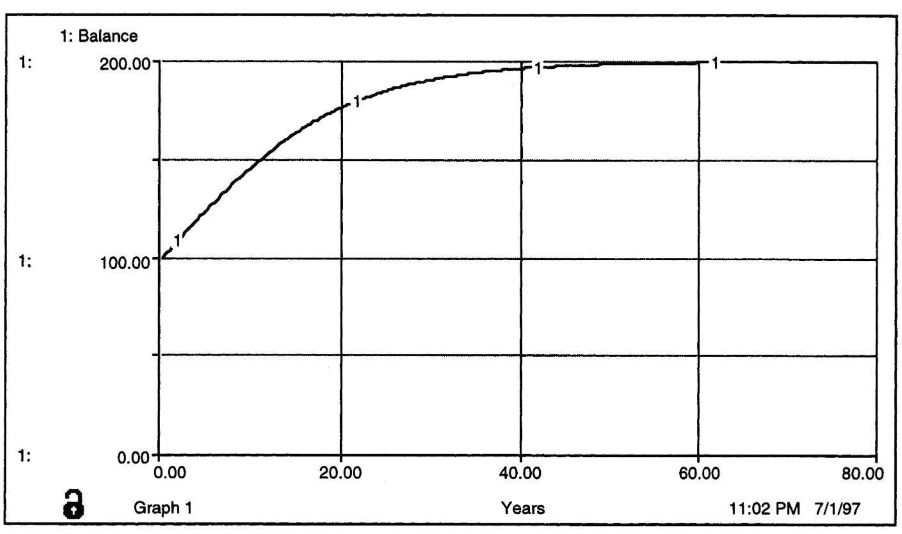


Figure 12

boxes corresponding to figure 12, we indicated that it is the *current* balance that determines the year-end payments and withdrawals. Such assumptions also underlie the Verhulst equation and the related models in which superlinear damping leads to sigmoid growth.

But what if a population manages to inject *delays* into the nonlinear damping term that determines the limits imposed by its carrying capacity? Such delays could arise as the result of a society's reliance on nonrenewable resources to sustain economic growth or the fact that it takes time for pollution to affect our health and degrade the soil.

Investigation 2 provides some important insights into such "delayed damping" in terms of a banking analogy. Here we suppose that Murky offers its established customers the following alternate way of computing the service fee on their accounts. After 10 years with Murky, your service fee will be based not on this year's balance, but on your balance 4, 6, or 8 years earlier. What is the effect of such delayed damping on money deposited with Murky?

STELLA provides an easy way of answering such questions. In programming figure 12's dialog box for fee, we simply replace **fee = 0.0005 · Balance · Balance** by **fee = DELAY(0.0005 · Balance · Balance, *d*)**, where *d* refers to a converter with a positive value. STELLA now enables us to do a "sensitivity run," plotting **Balance** for a range of values such as *d* = 4, 6, and 8. The outcome is shown in figure 13.

With this delayed damping model we are able to realize the four *behavior modes* that LTG associated with a growing population. We have already seen how positive feedback leads to *exponential growth* in the absence of limits and to *sigmoid growth* in the presence of superlinear damping with immediate feedback. The two other modes that STELLA has just helped us discover are called *overshoot and oscillation to equilibrium* and *overshoot and collapse.* In the case of our modified Verhulst equation, these two phenomena correspond to delayed feedback with small and large delays, respectively. In the context of World3, such phenomena are accompanied by an erosion of carrying capacity (also reflected in the Malthus–Condorcet model below). While far simpler than World3, investigation 2 does provide us with mathematical insight into LTG's final conclusion:

"If the world's people decide to strive for this second [sustainable] outcome rather than the first, the sooner they begin working to attain it, the greater will be their chances of success."

More questions and debate

Well, you have now seen some applications of STELLA and have gained some insight into dynamic modeling. But what about the implications of all this for *The Limits to Growth*?

The conclusion that I drew was that the framework that LTG provided for thinking about environmental issues was path breaking. First, it pioneered the use of computer technology to model environmental issues on a global scale in the context of a *closed system.* Second, it called for taking account of *delays* in building such models, an important innovation that theorists such as Verhulst were probably unable to handle. Indeed, investigation 2 in "Look, Ma—No Calculus!" corresponds to the numerical solution of a *delay differential equation* $dN/dt = 0.1N(t) - 0.0005N(t - d)^2$. The mathematical theory for such equations is scarcely 50 years old.

Others have drawn very different conclusions. In his recent book *How Many People Can the Earth Support?*, Joel Cohen worries about the extent to which a model such as World3 can truly reflect the complexity and unpredictability of the real world. How near to "the real thing" does a model have to be to play a useful role in debates on related social issues? Cohen refers to such models as "mathematical cartoons" that can be effective in both conveying and distorting truths.

However, Cohen himself is not immune from the lure of modeling. In a recent article entitled "Population Growth and Earth's Human Carrying Capacity" (*Science*, July 21, 1995), he develops a different generalization of the Verhulst equation, one that he calls the Malthus–

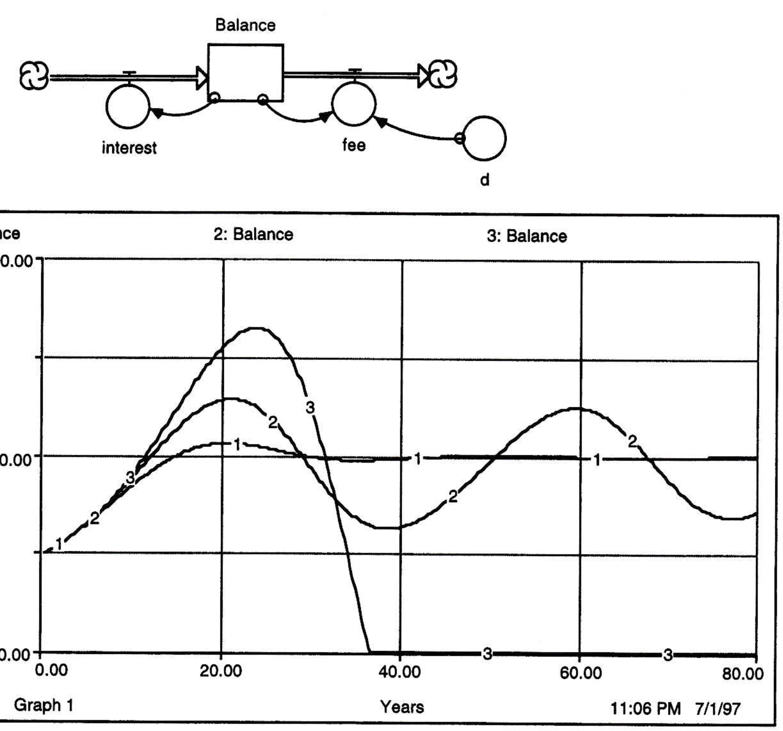


Figure 13

Condorcet model. Instead of relying on a single differential equation $dN/dt = aN(t) - bN(t)^2 = b[a/b - N(t)]N(t)$, Cohen considers a *system* of equations

$$\frac{dP}{dt} = rP(t)[K(t) - P(t)], \quad (2)$$

$$\frac{dK}{dt} = c \cdot \frac{dP}{dt}. \quad (3)$$

with the following interesting interpretation. Equation (2) resembles the Verhulst equation for a population of size $P(t)$ in an environment with carrying capacity K. However, as in the real world, K is not constant. Rather, according to equation (3), K has a growth rate that is proportional to the growth rate of $P(t)$. This reflects the point of view that "every human being represents hands to work and not just another mouth to feed." The question of whether the productivity of our hands exceeds the demands of our mouths would, in this model, be reflected by whether c is positive or negative. And it is the case $c < 0$ that corresponds to the erosion of resources, a concept that plays an important role in LTG.

In his analysis of these equations, Cohen observes that the "annual compounding" solution generates the same four behavior modes (exponential, sigmoid, overshoot and oscillation to equilibrium, and overshoot and collapse) encountered in LTG (and exhibited by our delay Verhulst equation $dN/dt = aN(t) - bN(t - d)^2$).

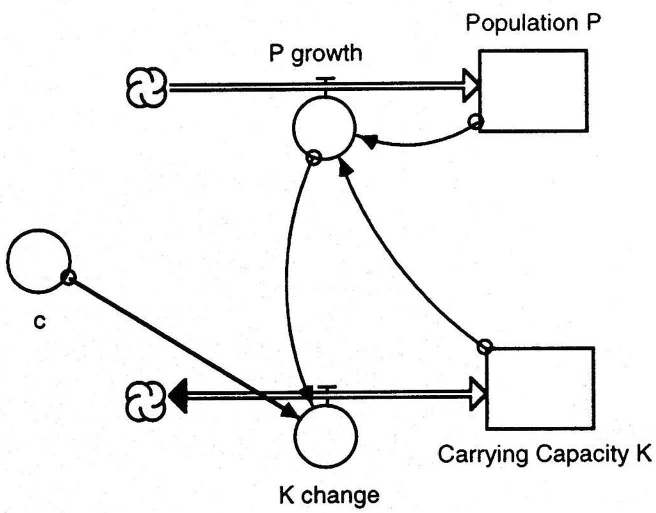


Figure 14

In his critique of *The Limits to Growth,* Cohen observes that a model as transparent as equations (2) and (3) suffices to generate behavior modes that LTG associated with World3. This leads him to ask whether the authors of LTG may have based their conclusions on an unnecessarily complex model, whose lack of transparency can mislead those unversed in system dynamics.

Rather than try to resolve such different interpretations, let us note that the Malthus–Condorcet model, recently cited in a premier research journal, is readily accessible to anyone equipped with systems software such as STELLA. The diagram in figure 14 will get you started.

In fact, you are now in a position to elaborate on Cohen's Malthus–Condorcet model as well. For example, you can use STELLA's built-in DELAY command to replace equation (2) with

$$\frac{dP}{dt} = rK(t)P(t) - rP(t-d)^2. \quad (2')$$

In this way computer technology and system dynamics can help you engage issues that lie at the cutting edge of scientific discourse.

Exercises

1. Suppose Murky tries to save money by updating your account every ten years, rather than every year. That means that after 10 years you would receive $10 \times 10\% = 100\%$ interest on your last balance and pay a service fee of $10 \times 0.05\% = 0.5\%$ on the square of your last balance. Calculate the value of $100 after 10, 20, and 30 years. What do you think would happen to $100 deposited at Murky for 100 years?

2. Suppose that Murky becomes even more sloppy, compounding every thirty years instead of every ten. Calculate the value of $100 after 30, 60, and 90 years. What do you think would happen to $100 deposited at Murky for 1,000 years?

3. When $c > 0$ in equation (3), a generalization might call for replacing c with $L/P(t)$, where L is a positive constant and $P(t)$ is the size of the population at time t. This corresponds to the assumption that a population's ability to enhance the Earth's carrying capacity will decline as P gets large. What additional connectors and converters are needed to accommodate this generalization in figure 14?

4. In the context of the Malthus–Condorcet model (equations (2) and (3)), generalize on exercise 3 so that $K(t)$ increases when $P < 2$ and decreases when $P > 2$.

Suggestions for further reading

D. L. Meadows et al., *The Limits to Growth* (New York: Universe Books, 1972)

J. Forrester, *World Dynamics* (Cambridge: Wright Allen Press, 1971)

Donella Meadows, Dennis Meadows, and Jørgen Randers, *Beyond the Limits* (White River Junction, Vermont: Chelsea Green Publishing, 1992)

D. L. Meadows et al., *Dynamics of Growth in a Finite World* (Portland, OR: Productivity Press, 1974)

D. L. Meadows et al., *Toward Global Equilibrium* (Portland, Oregon: Productivity Press, 1973)

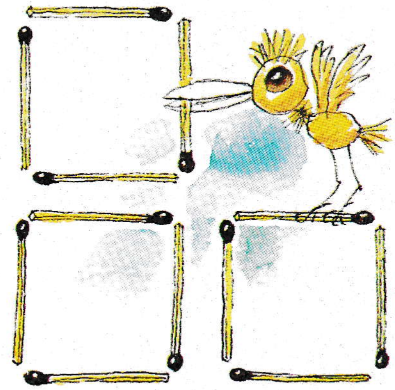
H. S. D. Cole et al., *Thinking About the Future: A Critique of The Limits to Growth* (London: Lamb, Chatto and Windus, 1973) [Also published as *Models of Doom* by Universe Books in 1973]

L. LaRouche, *There Are No Limits To Growth* (New York: New Benjamin Franklin House, 1983)

J. Cohen, *How Many People Can the Earth Support?* (New York: Norton, 1995)

Kurt Kreith *is a professor of mathematics at the University of California–Davis and is the editor of the LTG components of this special issue of* Quantum.

Just for the fun of it!



B211

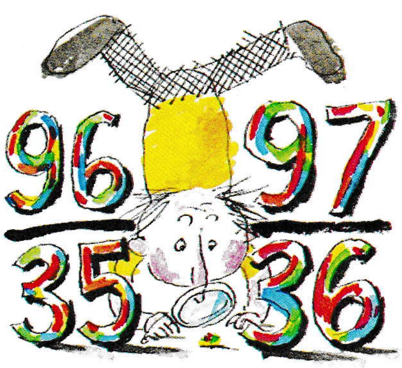
Match boxes. In the picture you see twelve matches arranged in three squares. Arrange the matches to form six squares (again with a side length of one whole match). (I. Sharygin)

B212

Six and three. Find six points on the plane such that each of them lies at a distance 1 from exactly three other points of this set. (I. Yaschenko)

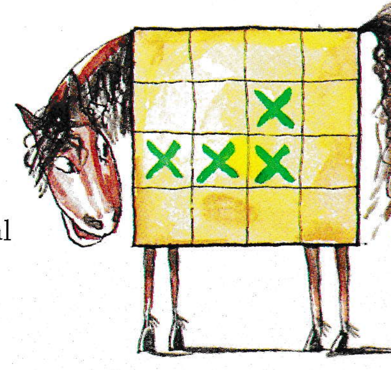
B213

Ratio of ratiocinators. Every seventh mathematician is a philosopher, and every ninth philosopher is a mathematician. So which is more numerous, mathematicians or philosophers? (A. Spivak)



B214

Look out below. Find the fraction between 96/35 and 97/36 with the smallest denominator. (D. Averyanov)



B215

Excellent square. Cut the square in the figure into four identical parts, so that each part contains exactly one X. (I. Sharygin)

ANSWERS, HINTS & SOLUTIONS ON PAGE 61

Art by Pavel Chernusky


more boats – more fish
more boats – less shade

Overshooting the limits

On plateaus and the various ways to reach them

by Bob Eberlein

T*HE LIMITS TO GROWTH* (LTG) HAS BEEN A topic of discussion for the last 25 years, and that discussion has often generated, as one of the authors Donella Meadows is fond of saying, more heat than light. Of the many points raised by this work, a very important one has often been overlooked because of its simplicity. While many have criticized *World Dynamics* and *The Limits to Growth* as being simply a restatement of Malthus's *Essay on the Principle of Population,* there is a fundamental difference. While Malthus tried to explain why the human condition was destined to be one of perpetual suffering, the LTG work addressed the dynamics of change, showing that the human condition could improve significantly, and degrade even more quickly.

Thomas Malthus was an English economist who lived from 1766 to 1834. He is most famous for his work suggesting that human population growth would proceed in unchecked exponential growth in the absence of any constraints. He also thought that food production could not ever do more than grow linearly and therefore concluded that most people would always live at a meager subsistence income level. It was this work of Malthus that led some to label economics "the dismal science."

Malthus relied on logic, mathematical analysis, and written arguments to support his position. *The Limits to Growth* made use of a simulation model. These two approaches do not represent exclusive alternatives—rather, they should be considered complementary. Mathematical and logical arguments can often be used to summarize the results of simulation studies. Simulation models can also be used to test conjectures and explore issues that do not easily lend themselves to rigorous mathematical analysis. In this article I will describe the use of simulation techniques to explore the so-called "overshoot problem."

Art by Sergey Ivanov

The computer simulation environment used in this article is Vensim®PLE (Personal Learning Edition). This software is free for educational use and information on how to obtain it is provided at the end of the article.

Population growth

At the heart of all this discussion is the fact that "a thousand millions are just as easily doubled every twenty-five years by the power of population as a thousand" (Malthus). In more mathematical terms, we would express this as "the rate of growth of population is proportional to the population itself." Using the stock and flow notation of system dynamics, we can represent this as shown in figure 1. Arrows with double lines are used to represent a pipe through which a "flow" leads to changes in a "stock." Births increase a population, and this is indicated by the arrow **births** going into **Population**. The variable **deaths** decreases **Population**, and this is shown by the right-pointing arrow. The clouds

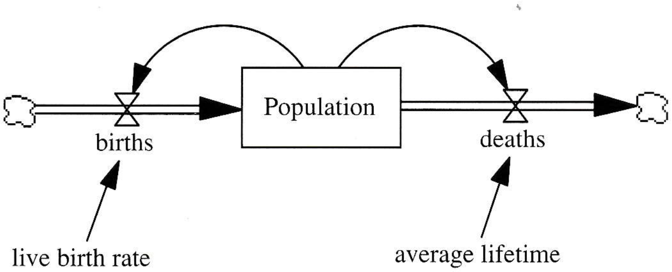


Figure 1
Simple population growth model.

shown at the ends of these arrows are used to indicate that source of **births** and repository of **deaths** are outside the boundaries of the model. The curved arrows that are shown connecting **Population** to **deaths**, **average lifetime** to **deaths**, and so on, are used to indicate functional dependence. That is, to determine the value for **deaths** we need to know both **Population** and **average lifetime**.

This picture represents a basic population growth model, but it is not complete. The picture says that **births** are determined by **Population** and **live birth rate**, but it does not specify the exact nature of that relationship. Behind the picture are the actual formulas or equations that specify the exact relationships between variables. For **births** we have

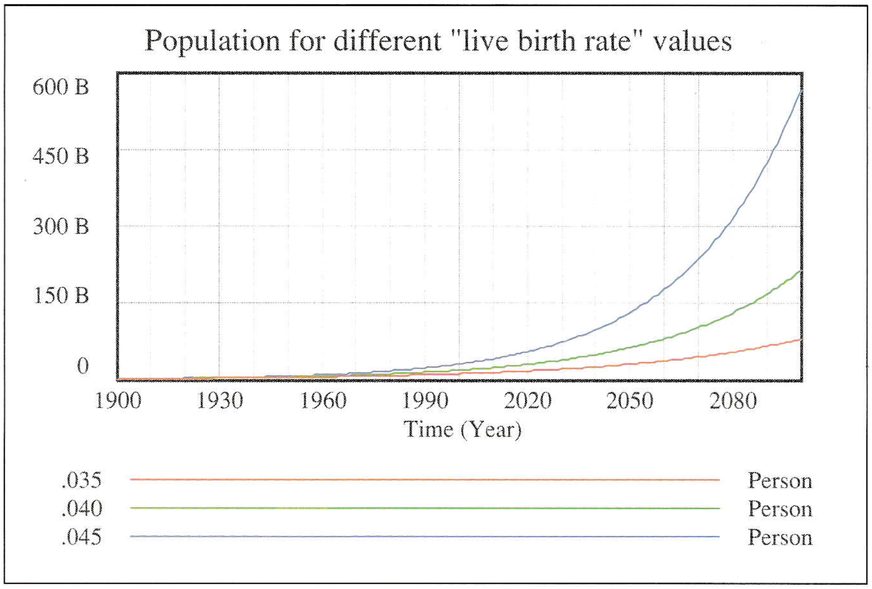


Figure 2

Unconstrained population growth.

births = Population * live birth rate (1)

This equation says that the number of **births** at any time is proportional to **Population** at that time. A unit of time is a central element to all system dynamics models, one that is often left implicit in the formulation of the underlying equations.

Our equation for **deaths** is formulated somewhat differently. Since a (stable) population with an average life span of 10 years has an annual death rate of 10%, we can substitute **1/average lifetime** for **death rate**. This leads to the equation

deaths = Population/average lifetime (2)

The equation for **Population** is a little different because **Population** is an accumulation—that is, a stock whose change in size is determined by **births minus deaths**. In equation (3) below we use Vensim notation to specify that an initial population of 1.65 billion is accumulating **births – deaths**—that is, adding each year's change to the previous year's value:

Population = INTEG(births – deaths, 1.65e9) (3)

For this simple model we could actually determine a closed-form solution that would give **Population** as a function of time (see problem 1 at the end of this article). This is, however, not possible in general, and the simulation approach to problems relies on numerical computation, not on finding "analytical solutions" in terms of formulas or algebraic expressions. Setting **average lifetime** to 65 years, we can simulate this model for several different values of **live birth rate** and see the results in figure 2.

For live birth rates larger than the death rate, all simulations generated by this model will exhibit the same type of behavior, namely exponential growth. The difference in results caused by seemingly small differences in **live birth rate** is striking.

Limited food supply

Malthus's most prominent thesis was that limitations in the food supply would prevent the unchecked growth in population that would otherwise occur. He argued that as the population rises, the amount of food per person falls, making disease, war, and famine more likely. In terms of our system dynamics approach, we would have to specify the functional relationships by which inadequate food supply contributes to deaths. This is contained in the **effect of food deaths function** in figure 3.

To specify the functional relations in figure 3, we replace equation (2) with

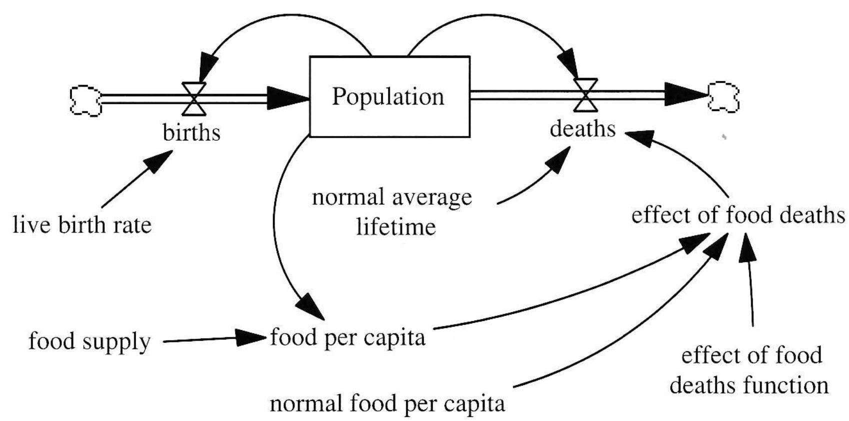


Figure 3

Population growth model with food constraint.

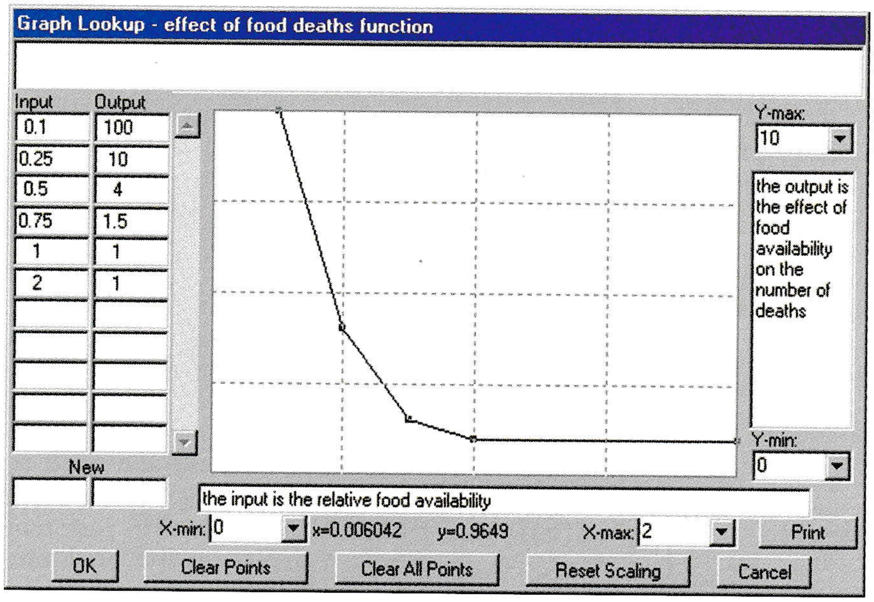


Figure 4

The "effect of food deaths" function.

deaths = (Population/normal average lifetime) * effect of food deaths (4)

In Vensim notation, the next equations is

effect of food deaths = effect of food deaths function(food per capita/normal food per capita) (5)

Measuring food in tons and food supply in tons/year, we now specify a nonlinear relationship between the ratio **food per capita/normal food per capita** and resulting increase in deaths. The Vensim notation

effect of food deaths function((0.1,100),(0.25,10),(0.5,4),(0.75,1.5),(1,1),(2,1)) (6)

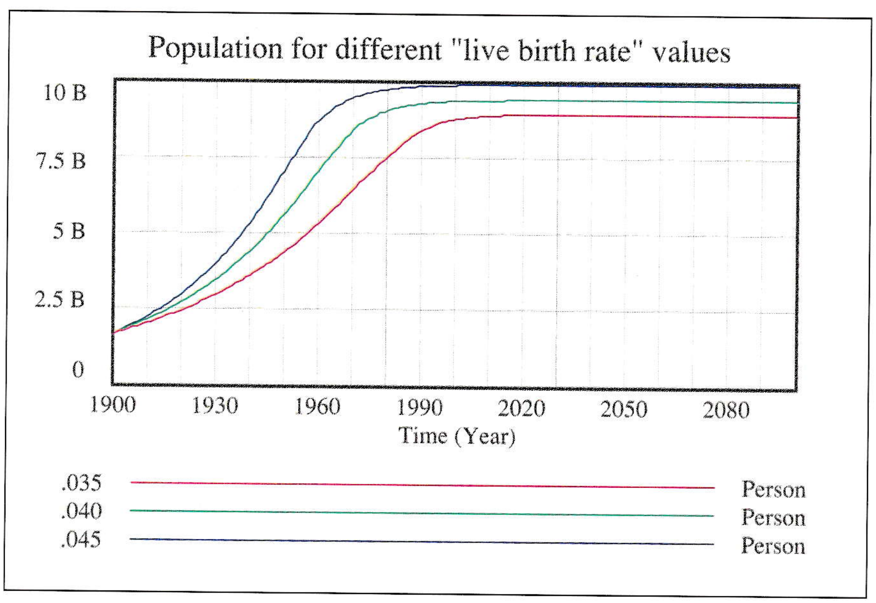


Figure 5

Population growth with a fixed food supply.

specifies a series of (x, y) pairs that gives rise to the function shown in figure 4.

In this context, we now complete the model with the following Vensim formulas:

food per capita = food supply/Population (7)

food supply = 3e+009 (8)

normal average lifetime = 65 (9)

normal food per capita = .5 (10)

The results of running such a simulation for several live birth rates are shown in figure 5. Three interesting conclusions can be drawn from this. First, the population grows and achieves a plateau just as Malthus said it would. Second, while there are differences that result from changes in the **live birth rate**, the population approaches an equilibrium regardless of this rate. Third, the lower the birth rate, the lower the equilibrium population reached and therefore the greater the equilibrium food per capita.

While our analysis has led to some interesting results, they are essentially the same as those of Malthus. However, the mechanisms we have modeled are more appropriate to a "hunter-gatherer" society than more complex social forms, including our modern industrial civilization. This leads us to ask, "What other things can happen?"

Nonrenewable resources

A fundamental difference between *The Limits to Growth* and figure 5 was the inclusion in LTG of a finite **Nonrenewable Resources** stock. This stock decreases as resources are used as part of economic activity. In order to relate this stock to the preceding model, we include a **target resource consumption** variable that is dependent on population. To acknowledge the dependence of agriculture on resources such as fuel, water, and fertilizer, we change the names of some variables and indicate the fact that a lack of resources can now also contribute to deaths (fig. 6 on the next page).

The top part of figure 6 is precisely the same as the one we saw previously, with **food** replaced by **resources** in the variable names. The variable **food supply** has been incorporated into a more complex relationship **resource consumption**, which is no longer a constant but is dependent on **target resource consumption** and **Nonrenewable Resources**. The Vensim equation for

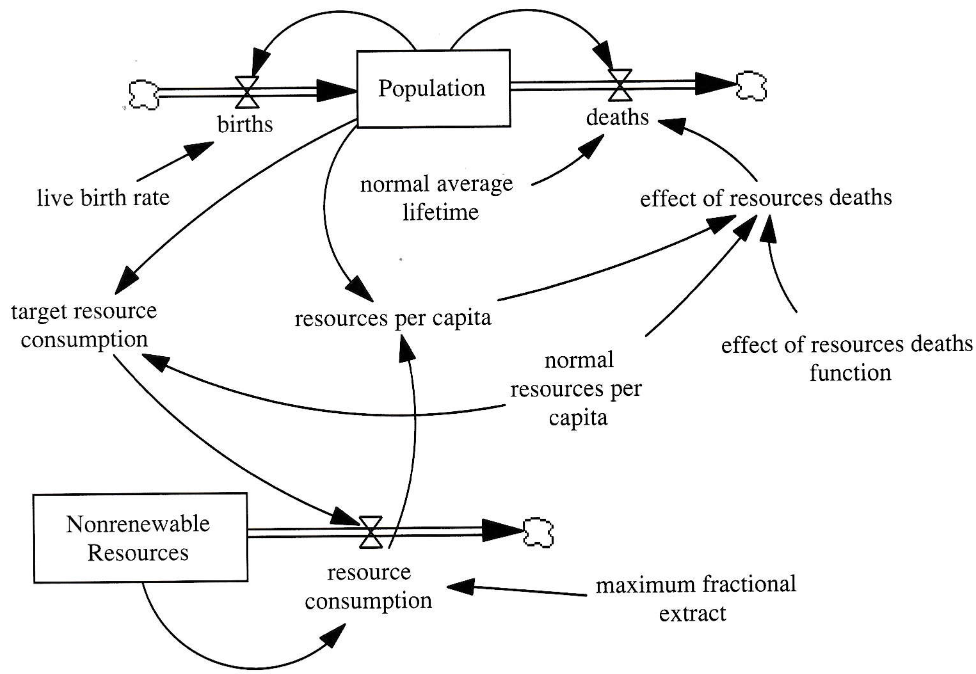


Figure 6

Population growth model with nonrenewable resources.

Nonrenewable Resource stocks will decrease exponentially, much like a radioactive substance with a half-life of about 140 years.

The results of running this simulation for various live birth rates are shown in figure 7. Here population is no longer increasing monotonically toward a plateau, but instead rises well beyond the levels achieved in preceding models and then falls. The different values for **live birth rate** change both the time and the population level at which the turnaround occurs.

By adding in the nonrenewable resource stock, we have introduced a new pattern of behavior for this model. This behavior results from the fact that **Nonrenewable Resources** are initially plentiful, but that their continued consumption eventually makes them insufficient to support an exponentially growing population. This result is not restricted to a society's dependence on nonrenewable resource.

Even if we allow for resource renewal at a constant rate, an overshoot of limits still occurs. The underlying problem is that there is an endowment of resources that can support a very large population for a time, but that expo-

resource consumption is

resource consumption = MIN(target resource consumption, Nonrenewable Resources * maximum fractional extraction). (11)

This equation says that resources are consumed at a rate to meet the demands of the population except that there is a limit on how fast the remaining resources can be extracted.

The equation for **target resource consumption** is

target resource consumption = Population * normal resources per capita (12)

and the equation for **Nonrenewable Resources** is

Nonrenewable Resources = INTEG(–resource consumption,Se+12) (13)

This last equation indicates that an initial resource stock of $5 \cdot 10^{12}$ units is decreased by **resource consumption**. The resource units correspond to an amount that would provide for ten billion people for 1,000 years at the normal consumption rate of 10 units/person/year. The **maximum fractional extraction** is set at 0.005/year, which corresponds to an accelerated extraction time of 200 years. It can be shown (see problem 2 below) that once the population is sufficiently large to put pressure on resource extraction, the

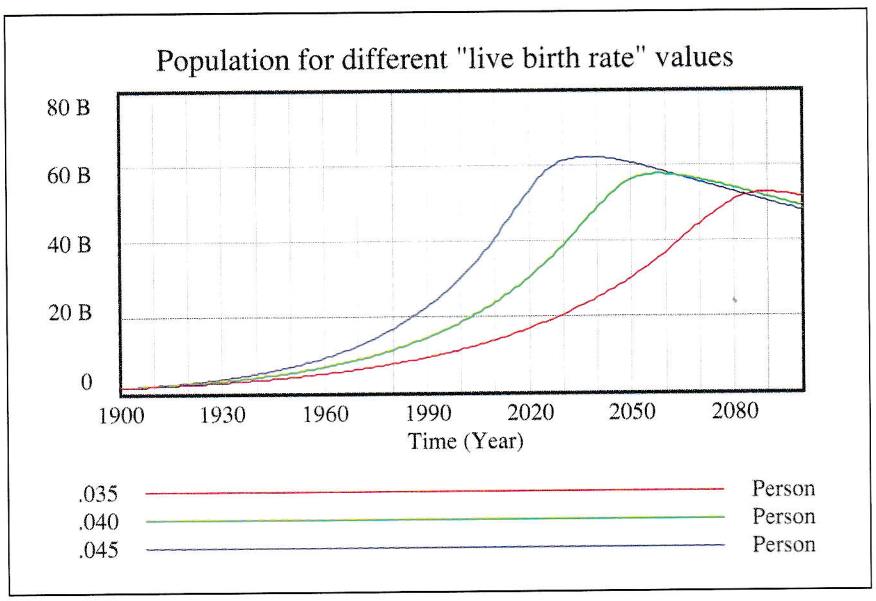


Figure 7

Population growth with nonrenewable resources.

About Vensim® PLE

Vensim PLE is one member of the Vensim family of software designed to make it easier to develop and use high quality system dynamics models. Vensim PLE is free for educational and personal use. It can be downloaded from the World Wide Web at http://www.vensim.com. Vensim PLE includes both the World2 model documented in *World Dynamics* and the World3 model as updated for *Beyond the Limits.* For more information contact

Ventana Systems, Inc.
149 Waverley Street
Belmont MA 02178 USA

Phone: 617 489 5249
Fax: 617 489 5316
E-mail: vensim@vensim.com
URL: http://www.vensim.com

nential growth, left unabated, will exhaust any resource that is not itself growing exponentially. A tragic example is that of Easter Island. Here islanders harvested trees faster than they could replenish themselves and, as a result, were unable to build canoes from which to fish.

In equation (11) above we specified that resources are to be consumed according to the population's needs, except that no more that 1/2 of one percent of the existing resources can be consumed in a single year. This is a fairly gentle way of bumping into a resource limit. A more complex and perhaps realistic formulation, such as that used in the World models, can lead to more abrupt overshoot and decline.

Conclusions

The lessons of *The Limits To Growth* go far beyond what we have touched on here. The centrally important point made in this article is that adjustment to limits need not involve a monotonic approach to a plateau. As more realism is introduced into the model, taking into account the interactions among pollution, industrial capital, agriculture, and land, the overshoot and subsequent decline become significantly more pronounced.

Computer simulation provides an engaging way to address the monotonicity question and convey it without reference to calculus and differential equations. As a building block for investigation, simulation has another big advantage. While it may be tough to start out with an elaborate structure for a model, simulation enables us to build on simpler structures. Furthermore, once additional structure is included, the resulting model still allows for solutions based on simulation. By contrast, only the simplest problems allow for closed-form solutions, and this can make heuristic insights very difficult to develop.

Experimentation with simulation will always give results. Intelligently used, these results can provide important new insights and a great deal of learning.

Problems

1. An annual census reveals that an population of 100 rabbits has a live birth rate of 20%/year and an average life span of 10 years. Find a formula for the number of rabbits at the end of n years.

2. A population uses $^{1}/_{200}$th of its remaining nonrenewable resources each year. How many years will it take for its resource reserves to be reduced to 50% of their original level?

Bob Eberlein *is the head of product development at Ventana Systems, Inc. He holds a Ph.D. in Management from MIT and has been consulting and teaching in the field of system dynamics since 1984.*


Can you hold your breath?
It's your turn!

The world in a bubble

"Planetary management" in the closed ecological systems of Biosphere 2

by Joshua L. Tosteson

IN 1968, AN ECLECTIC INTERnational group of scientists, economists, politicians, and educators gave birth to a new organization, the Club of Rome, which soon thereafter commissioned a groundbreaking study known as the Project on the Predicament of Mankind. The goal of the project was to address humanity's growing impacts upon the global system, and the predicament that "despite (our) considerable knowledge and skills, (we do) not understand the origins, significance, and interrelationships of its many components and thus (are) unable to devise effective responses." *The Limits to Growth* (LTG for short), whose legacy and influence we explore in this issue of *Quantum,* was the project's effort—and perhaps humanity's first systematic attempt—to map out and assess these interrelationships on a planetary scale.

Perhaps LTG's most important contribution to our intellectual heritage lies in the way that it forced us to think about the Earth's biosphere—the interconnected natural systems within which human economies and societies are situated, that enable life on the planet to flourish—as a "closed" system. In contrast to "open" systems, in which matter can move in and out, closed systems do not permit the transfer of materials. For most of human history people have acted as if the Earth is an open system, with limitless resources and an infinite capacity to absorb human impacts. This was an easy assumption to make because the scale of human civilization, relative to the size of planet, had been small up until the past one hundred and fifty years.

But with an ever growing, technologically advancing human population, the scale of human activity on the Earth since the mid-1800s has made us acutely aware that the Earth's resources and life support systems do indeed have limits. While the Earth's biosphere is energetically and informationally open (meaning that sunlight, gravity, and other sources of energy and force act upon the Earth), it is in fact a materially closed system—sort of like a huge, sealed jar. Except for the infrequent escape of light elements into space and the occasional intrusion of a meteor or asteroid, the Earth does not gain or lose any matter. LTG powerfully showed us, through the computer model World3, how human beings could conceivably exhaust the Earth's supply of natural resources and test the limits of the planet's ability to support humans. Despite the model's power, however, it was still difficult for people to picture how its results really connected to their lives and to the real world around them. Twenty-five years later a living model, a human experiment inside a sealed miniworld, has again illuminated many of the themes that LTG first brought to the world's attention, in a particularly immediate and graphic way.

Covering 3.15 acres of area in the high Sonoran desert just north of Tucson, Arizona, Biosphere 2 is a research facility that has captivated the imagination of both scientists and the public throughout its occasionally controversial six years of life. The facility houses five wilderness biomes (rainforest, desert, savanna and thornscrub, estuary, and ocean with coral reef), an intensive agriculture biome (IAB), and human living quarters (see figure 1 on the next page). From 1991 to 1994, the facility was used to support two live-in crews within the enclosure, during which time the crews remained sealed inside Biosphere 2 and were responsible for meeting all of their

Art by Sergey Ivanov

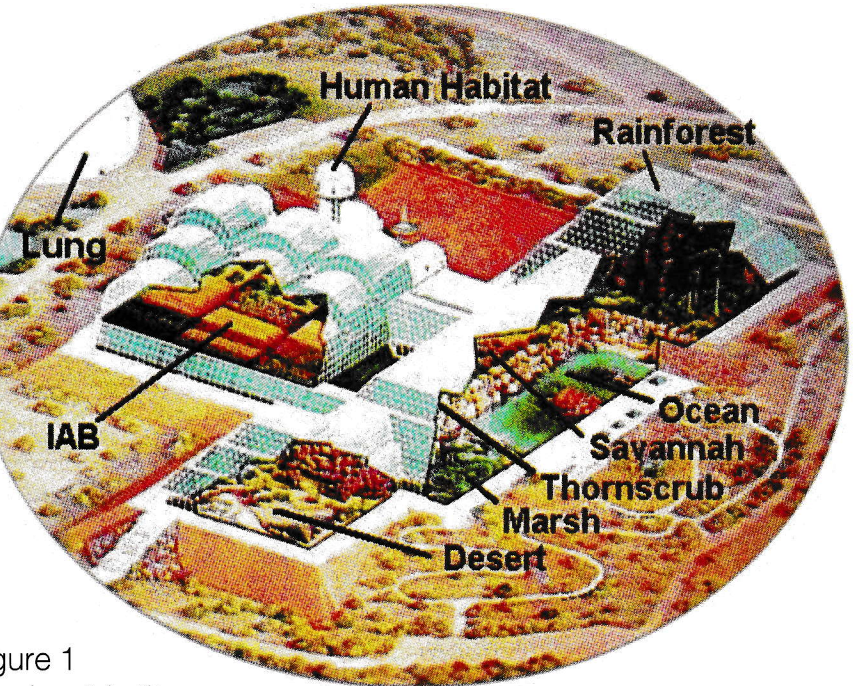


Figure 1
Biosphere 2 facility.

survival needs. More recently under the management of Columbia University, the facility has been utilized to investigate scientific questions pertinent to human impacts upon the global environment.

As a materially closed system, Biosphere 2's climate, atmospheric composition, recycling of water and nutrients— all of the basic life support functions that the Earth provides for us—must be almost completely engineered by humans. To support human life within the enclosure, the original project scientists had to manage the environmental conditions of Biosphere 2's biomes to maximize food production, maintain a safe balance between oxygen and carbon dioxide levels, preserve a high level of biodiversity, and successfully recycle water and nutrients for drinking and rainwater—all in the context of material closure. In short, they had to manage all of the basic life-support functions of the planet within an enclosure not much bigger than a couple of football fields!

In this article I will look at some of the challenges that we have had in managing the complex closed system of Biosphere 2. I'll focus mostly on the story of Mission One in Biosphere 2, when eight crew members lived inside the Biosphere for two years, from 1991 to 1993. Then I'll take some of the lessons that were learned from that experience to reflect on the Earth ("Biosphere 1") as a closed system and to take a fresh look at some of the ideas that were born twenty-five years ago in *The Limits to Growth*. Along the way, I'll present some problems that are intended to stimulate further exploration of the themes that are touched on in the article.

Oxygen loss, concrete, and the microbial feast: Mission One in Biosphere 2

What would our lives be like if we had to manage the Earth in order to maintain oxygen levels in the atmosphere at safe levels (between 19% and 21%)? What if international emergency task forces had to be deployed to find ways of keeping atmospheric carbon dioxide (CO_2) levels from climbing above 2,000 parts per million (current CO_2 levels are about 355 parts per million)? What if small mistakes in the strategies chosen to deal with these problems could send the planet into a catastrophic spiral? What if, in short, we had to engineer the Earth's life support systems to meet all of our survival needs?

This is, in a nutshell, what life for the "Biospherian" crews was like. For two years, from September 1991 to September 1993, eight individuals lived inside the enclosed facility of Biosphere 2 (a second crew also lived inside the facility between March and September 1994). During this period, the crew members' priorities were to feed themselves, make sure that the atmosphere's chemical composition remained safe, and maintain a high level of biodiversity. They operated under the assumption that the system would remain closed for 100 years and that no new species would be introduced—in other words, that the Biosphere would be operated as a completely closed system. As it turned out, their goals were almost impossible to achieve, for reasons we will explore below.

Rainforest management

To begin, let's look at how the crew managed the rainforest biome to help meet their survival needs. The rainforest was a critically important biome for the crew, for a number of reasons. First, the rainforest held a great deal of the Biosphere's biodiversity, and the crew was eager to preserve it. But more importantly from a survival point of view, the rainforest also played a huge role in helping to maintain the balance of oxygen and CO_2 in the Biosphere's atmosphere. Plants "breathe," or sequester, carbon dioxide through the process of photosynthesis, while exhaling oxygen at the same time. This process is described by the following reaction:

$$CO_2 + H_2O \rightarrow CH_2O + O_2, \quad (1)$$

where CH_2O is a general formula for organic material made by the plant after assimilating CO_2 from the atmosphere.

Because rainforests are extremely productive ecosystems (meaning that their plants take in carbon dioxide at very fast rates), they produce

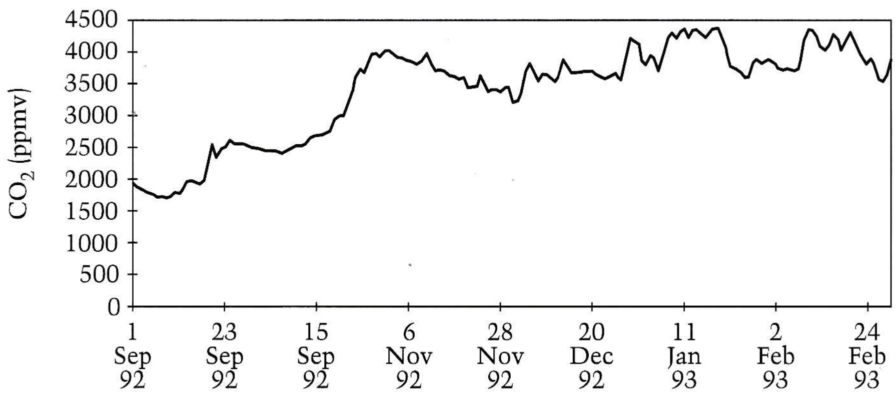


Figure 2

CO_2 levels in Biosphere 2, September 1992 to March 1993.

much of the oxygen on Earth (Biosphere 1) and help keep carbon dioxide levels relatively low in the Earth's atmosphere. In Biosphere 2, the rainforest biome served a similar function: to ensure that oxygen levels stayed high (around 21%) and CO_2 levels stayed relatively low.

But by the fall of 1992, one year after the Biosphere was initially sealed, the Biospherians became concerned about the rising levels of CO_2 in the facility (fig. 2), as well as the steady decrease in oxygen levels (fig. 3). Clearly photosynthesis was not able to keep pace with some other process that was adding CO_2 to, and taking O_2 out of, the atmosphere. In this context, the role of the rainforest in helping the Biospherians boost oxygen levels and stop the rise of carbon dioxide became urgently important.

The Biosphere's rainforest is approximately 1,900 m^2 and has a volume of 35,000 m^3—about 1/4 the total volume of the enclosure (see figure 1). It is 22 m from the ground at its highest point. Air handlers, located in the basement underneath all of the biomes, produce flows of chilled and heated air through the rainforest, thereby controlling temperature and humidity. Ground and overhead sprinklers produce rain, and a fogging system enables the rainforest to be run at very high humidity levels. Sensors within the rainforest monitor atmospheric composition, temperature, light, and humidity, which allowed the Biospherians and the project scientists outside the facility to monitor the conditions of the Biosphere in real-time. (You, too, can look at Biosphere 2 data in real time, by visiting the Biosphere 2 World Wide Web site at www.bio2.edu).

With these resources at their disposal, the crew encouraged the growth of fast-growing weedy species because of their ability to take up carbon and produce oxygen rapidly. Elsewhere in the rainforest, the crew pruned plants to stimulate more photosynthesis. But keep in mind that the Biosphere is a closed system. The crew could not just return the pruned plant material to the soil, because it would decompose and the carbon that the plants took out of the atmosphere (which they stored in their tissue) would be released back into the atmosphere. Microbes—small bacteria that live in the soil—use organic material (dead plants and animals) as their source of food. Their decomposition of soil carbon pulls oxygen out of the atmosphere and releases carbon dioxide back into the atmosphere. This process is described by the following reaction:

$$O_2 + CH_2O_4 \rightarrow CO_2 + H_2O, \quad (2)$$

where CH_2O represents a general formula for organic material in the soil.

So the crew had to store the pruned biomass in the basement to keep the carbon that it took out of the atmosphere from returning back into active circulation within the system. In this way, the Biospherians hoped to set up a process by which carbon was consistently removed from the atmosphere, while oxygen was continually added to the atmosphere. Living in a closed system was not easy!

Problems

The volatility of a system's carbon cycle (the biological, soil, and atmospheric reservoirs that hold carbon in a system and the processes that transfer carbon among them) is closely related to the time it would take for the flows to "flush out" all CO_2 or CH_2O from its carbon reservoirs. If the system is in equilibrium (inflow equals outflow), this index of volatility is

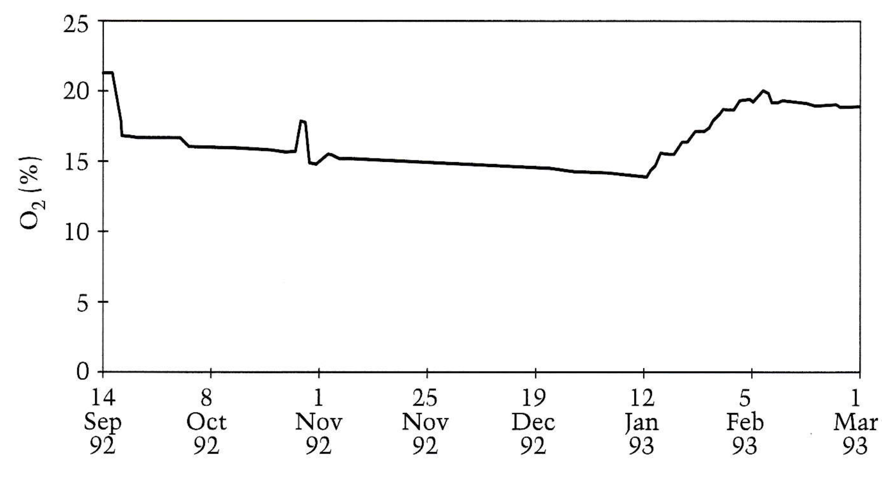


Figure 3

O_2 levels in Biosphere 2, September 1992 to March 1993.

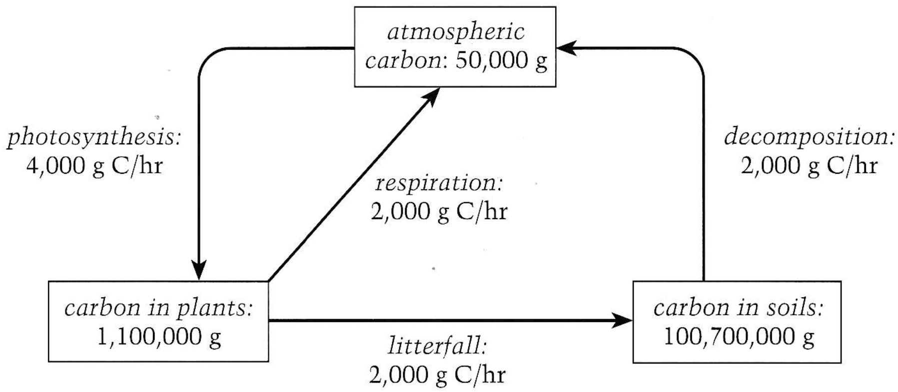


Figure 4

Simple model of the carbon cycle inside the Biosphere 2 rainforest biome.

obtained by dividing the amount of carbon held in a reservoir by the size of the inflow (or outflow) of carbon. By assuming that all carbon atoms spend the same amount of time in a given reservoir, this index can also be thought of as the "residence time" of carbon-bearing molecules.

Problem 1. Figure 4 represents a simplified equilibrium model for the carbon cycle in Biosphere 2's rainforest biome. Here we measure carbon in grams and carbon flows in grams/hour. Calculate the residence times for carbon-bearing molecules in the rainforest's atmosphere, plants, and soils.

Problem 2. Figure 5 represents a simplified equilibrium model for the Earth's carbon cycle prior to the industrial revolution. Here we measure carbon in gigatons (Gt) and carbon flows in Gt/year. Calculate the residence times of carbon-bearing molecules in the Earth's pre-industrial atmosphere, plants, soils, and oceans.

Problem 3. It is estimated that prior to the Industrial Revolution, the Earth's atmosphere contained about 0.028% CO_2, or 280 parts per million (ppm). Data gathered at the Mauna Loa Observatory in Hawaii (fig. 6) indicates that from 1960 to 1990, the CO_2 level in the Earth's atmosphere rose at about 2 ppm/year. Assuming this additional CO_2 represents the burning of fossil fuels, estimate the number of gigatons of carbon burned each year.

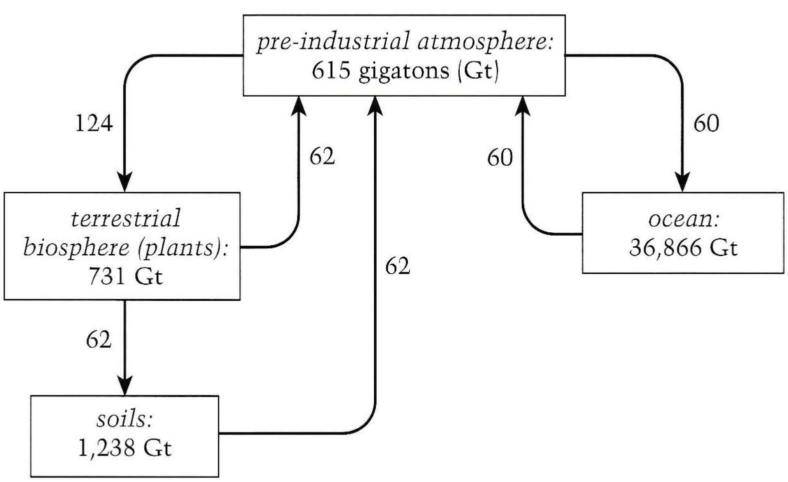


Figure 5

Global carbon cycle (after McElroy, "The Atmosphere: An Essential Component of the Planet's Life Support System," text for Science A-30, Harvard University).

Where's the carbon?

Despite their tireless efforts, the crew's strategies for managing the rainforest did little to halt the loss of oxygen from the atmosphere, as figure 3 clearly shows. In fact, oxygen levels dipped so precipitously low in January 1993—to a health-threatening 14%—that the project officials made the difficult decision to pump oxygen into the Biosphere via the "lungs" (hence the sharp increase in oxygen levels depicted in figure 3). Fortunately for the crew, the Biosphere could be made an open system in an emergency. Had the Biosphere been in space, the crew would surely have died.

Former Biospherian crew member Linda Leigh recalled her tiring journey through the Biosphere to the lungs to get a first whiff of injected oxygen, reflecting that there was "something very poetic about taking an expedition to a lung in order to breathe." It was clear by then that the crew's strategy was ineffective at meeting their survival needs. What no one knew at the time, however, was why this was the case.

Researchers from Columbia University were called in to try and find out what was going on. They quickly determined that the cause of the problem likely lay within the carbon-rich soils of Biosphere 2. The soils of Biosphere 2 were originally loaded with a large amount of organic matter, which (as we learned earlier) microbes in the soil use as their source of food. Because there was so much organic carbon in Biosphere 2, they reasoned, the process of decomposition occurred at an extremely rapid rate. And because the glass and spaceframe of the facility blocked as much as 45% of the external light from getting to the plants, the photosynthesis rate of the plants could not keep up with the decomposition rate in the soil. This led, they hypothesized, to the constant buildup of CO_2 and depletion of O_2 in the Biosphere 2 atmosphere.

For this hypothesis to be valid, however, the number of CO_2 mol-

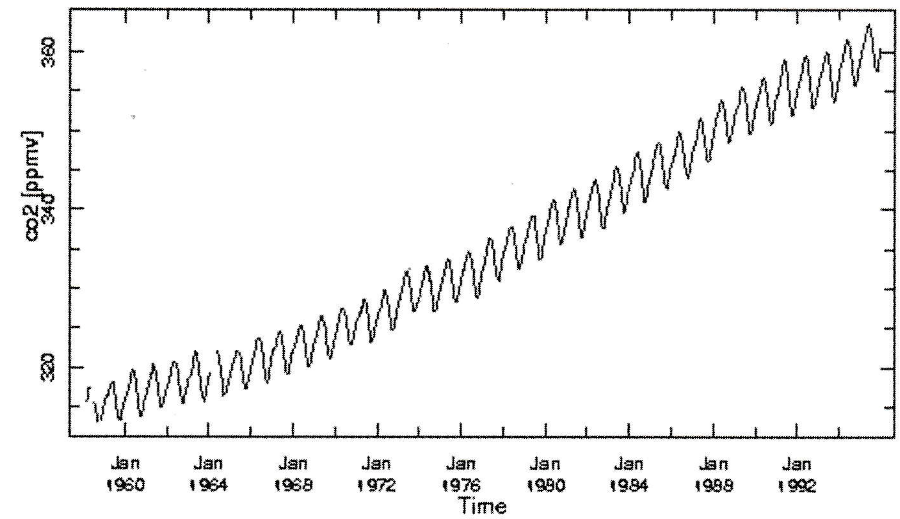


Figure 6

Atmospheric CO_2 levels measured over Mauna Loa, Hawaii (source: Carbon Dioxide Information Analysis Center).

ecules (expressed in moles) added to the Biosphere 2 atmosphere should have been equal to number of O_2 molecules lost from the atmosphere. This is because, as equations (1) and (2) show, for every CO_2 molecule taken up or released by photosynthesis or decomposition, an O_2 molecule is also released or taken up.

The data in figure 7, however, reveal that the number of moles of CO_2 that were released to the Biosphere 2 atmosphere was less that the number of moles of O_2 that were taken out between September 1992 and January 1993. Over that period, only about $1.1 \cdot 10^4$ moles of CO_2 were added to the atmosphere, while about $5.4 \cdot 10^5$ moles of O_2 were taken out.

Clearly the total amount of CO_2 released to the atmosphere was quite a bit less than the total amount of O_2 taken out from the atmosphere—a full order of magnitude, in fact. However, despite appearances to the contrary, these data do not necessarily invalidate the research team's original hypothesis. Why? Because there could have been some process in Biosphere 2 occurring independently of the ecosystem processes we have examined (photosynthesis and decomposition) that was taking carbon out of the atmosphere without releasing oxygen back into the atmosphere. As it turned out, two such processes were indeed at work.

Gas measured, date	Measurement
CO_2, September 1992	2,000 ppm
CO_2, January 1993	4,000 ppm
O_2, September 1992	21%
O_2, January 1993	14%

Figure 7

Changes in carbon dioxide and oxygen levels in Biosphere 2.

Answers in unlikely places

First, chemical scrubbers had been automatically pulling CO_2 out of the atmosphere for quite some time. But quick calculations revealed that the amount of carbon dioxide "scrubbed" from the atmosphere could not come close to accounting for the total amount of "missing" carbon.

Puzzled, the team began to investigate a number of possibilities to try and address the problem. At one point, it was proposed that the concrete in Biosphere 2 had been reacting with the CO_2 in the air via the following reaction:

$$CO_2 + Ca(OH) \rightarrow CaCO_3 + H_2O. \quad (3)$$

Suspecting this was the case, they took samples from the concrete. If the concrete were indeed reacting with atmospheric CO_2, then it should have contained a high abundance of $CaCO_3$ (carbonates) They found that, indeed, high levels of $CaCO_3$ were present in the concrete, enough to confirm that the concrete had removed sufficient atmospheric carbon to hide the effects of the rapid decomposition and slow photosynthesis. So, as it turned out, the chemical scrubber and, more importantly, the concrete masked the most important dynamic of the system for the health of its inhabitants—namely, that decomposition significantly outpaced photosynthesis because of the overabundance of organic material in the Biosphere 2 soils.

While inside the enclosure, the Biospherians' biome management strategy seemed to be an intelligent way to deal with the problem of maintaining a healthy atmospheric composition, while still preserving a high level of biodiversity. But because they lacked appropriate, timely knowledge of a very complex system, the Biospherians could not know why their strategy was ineffective—they could only observe the ever decreasing oxygen levels inside Biosphere 2. Given the initial conditions of the Biosphere (high soil organic carbon and low light levels for two years), how might you have managed the system—without any inputs from outside the system—to mitigate the buildup of CO_2 and the loss of oxygen? Do you think that the Biospherians' strategy was well thought out?

A tale of two closed systems

Despite their best efforts, the Biospherian crew members were unable to find a solution to their problems in a way that enabled them to retain Biosphere 2's material closure. It was lucky indeed that they were able to add oxygen at a critical juncture, or they would surely have met an unhappy end. Their almost tragic story offers us a contemporary allegory, a cautionary tale for us as we approach the 21st century.

First and foremost, the Biospherians' experience reminds us that we do not know nearly enough about how our planet works to build even a very tiny, miniature version of it that would capture all of its most important features. In this sense Biosphere 2, as a model, failed to reproduce the essential behaviors of the larger system (the Earth) that it was built to simulate. The experience of Mission One reveals that our planet is an unimaginably complex, closed system that will always continue to surprise us with its behavior—happily in some cases and perhaps tragically in others (as in the Biospherians' brush with oxygen starvation).

Despite our lack of knowledge and the inevitability of surprises, however, we know what can happen when a really complicated system gets thrown out of equilibrium. In Biosphere 2, the system designers made a fundamental error. They set up the plant–soil–atmosphere system in such a way that the balance of CO_2 and O_2, which is established by photosynthesis and decomposition rates, could not be maintained in a way that supported human life for very long. In Biosphere 1, we have now become the system designers, whether we are fully aware of it or not. By altering the reservoirs and flows of countless chemical species and natural resources, from plants and soils to minerals and waters, we are in effect redesigning an extremely complex system that has been in a dynamic equilibrium for millennia. And we are receiving a variety of indicators from the Earth system—the loss of stratospheric ozone from CFCs, the rise in atmospheric CO_2 levels from fossil fuel combustion, and many others—that bear striking resemblance to the frightening loss of O_2 that the Biospherians witnessed, but were simply unable to stop in a relevant time frame. Through a flawed design, some mistakes, and plain bad luck, the Biospherians reached and exceeded the limits of the Biosphere's ability to support people.

The question raised by all of this is: how close are we to the limits of the Earth's ability to support people? Are we perturbing the Earth's systems in such a way that they may no longer be able to perform their life support functions for us? As Biosphere 2 showed us, we are not very good at building or managing planets. But, unlike in Biosphere 2, we don't have an airlock door that we can walk through if things go wrong.

Acknowledgments

The author wishes to thank Dr. Kurt Kreith for his many helpful comments and advice on the article. Dr. Bruno Marino also provided useful commentary, as did Dr. Debra Colodner. Most of this article was written under the generous support of Biosphere 2 Center, an affiliate of Columbia University. The author gratefully acknowledges this support.

ANSWERS, HINTS & SOLUTIONS ON PAGE 62

Joshua L. Tosteson *has been involved with the Biosphere 2 Center (located in Oracle, Arizona) since 1994, first as a Fellow in Environmental Science, Public Policy, and Education, and later as Curriculum Coordinator for the center's education programs. He is currently pursuing doctoral studies at Columbia University's Earth Institute and Department of Earth and Environmental Sciences.*

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M211

Plane coincidence. Find all pairs of numbers m and n such that the set of points on the plane whose coordinates satisfy the equation $|y - 2x| = x$ coincides with the set defined by the equation $|mx + ny| = y$. (D. Averyanov)

M212

Meeting of circumscriptions. In triangle ABC sides CB and CA are equal to a and b, respectively. The bisector of the angle ACB intersects side AB at point K, and the circle circumscribed about the triangle intersects it at point M. The second point at which the circle circumscribed about the triangle AMK meets line CA is P. Find the length of AP. (V. Protasov)

M213

A bracing problem. Solve the following system for arbitrary a, b and c:

$$\begin{cases} \dfrac{a}{x} - \dfrac{b}{z} + xz = c, \\ \dfrac{b}{y} - \dfrac{c}{x} + xy = a, \\ \dfrac{c}{z} - \dfrac{a}{y} + yz = b. \end{cases}$$

(I. Sharygin)

M214

Tetrahedral conditions. Prove that one can fold a given paper triangle so that it covers without overlap the surface of a unit regular tetrahedron (that is, a triangular pyramid whose edges are all equal to 1), if (a) the triangle is isosceles, with legs of length 2 and vertex angle equal to 120°; (b) two sides of the triangle are equal to 2 and $2\sqrt{3}$ and the angle between them is 150°. (I. Sharygin)

M215

Let 1997 be. Let $b(n)$ denote the number of ways of representing n in the form

$$n = a_0 + a_1 \cdot 2 + a_2 \cdot 2^2 + \ldots + a_k \cdot 2^k,$$

where the coefficients a_i, $i = 0, 1, 2, \ldots, k$, can be equal to 0, 1, or 2. Find $b(1997)$. (V. Protasov)

Physics

P211

Athletes bound. Two runners joined by an elastic cord are standing at points A and B. They start to run simultaneously: runner A to the east with a velocity $v_0 = 1$ m/s, and runner B to the south with some constant acceleration. Find this acceleration if it is known that a knot C tied on the cord passes through a given point D (see figure 1). (S. Krotov)

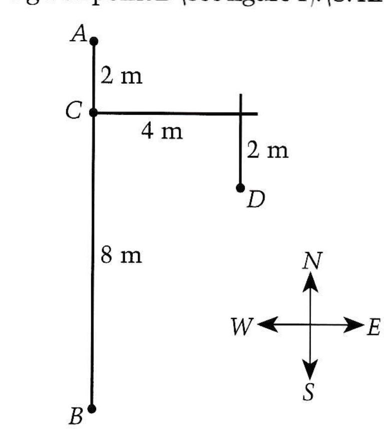


Figure 1

P212

Wind tunnel warmth. A model of a dirigible is tested in a wind tunnel,

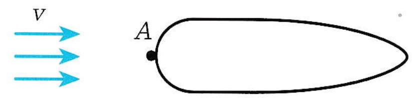


Figure 2

where an air flow of speed $v = 300$ m/s is directed at it. At point A (located precisely on its axis) the speed of the flow drops to zero (see figure 2). Find the temperature of the air near this point. The temperature of the surrounding air is $T = 300$ K. (A. Zilberman)

P213

Mercurial shakedown. Why does it take so long to take a person's temperature with a mercury thermometer (about 10 minutes), whereas one can shake the mercury back down almost immediately after the measurement is taken? (G. Kosourov)

P214

Splintered charge. A conducting sphere exploded and produced a number of fragments, which scattered over a large distance. The splinters are arbitrarily connected by thin wires. Which is larger: the electric capacitance of the system of splinters or that of the original sphere? Neglect the capacitance of the wires. (F. Lutsenko)

P215

Reflection of a sunbeam. An observer catches a sunbeam in a small mirror while standing in front of a large one in which she sees her image. What will she see if she directs the sunbeam at the image of the small mirror in the large one? (S. Krotov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 59


you'll go there anyway...
when I die I'll go to the clouds...

Learning from a virus

An application of systems thinking and dynamic modeling

by Matthias Ruth

DID YOU EVER WONDER WHY, AFTER A LARGE number of people suffer from the flu, the epidemic seemingly disappears, only to reappear again a few weeks later? Or did you ever wonder how it happens that economists insistently talk about the equilibrium of demand and supply, yet prices tend to fluctuate and never really settle down? And how is it possible that over many years, forests in Canada and the United States seem lush and healthy, and then, within one summer, insect outbreaks decimate the foliage and turn the forest from green to brown? The insects are virtually gone by the next year, and the forest slowly returns to its original state, only to be infected again a few years later.

Nonlinear dynamic systems

These phenomena have a number of features in common. First, the systems that generate such seemingly erratic behavior consist of individual parts that interact with each other. In the case of influenza, there is one group of people who carry the virus and a second group that receives the virus upon contact with members of the first. In the case of demand and supply interactions, we have a marketplace within which producers and consumers exchange goods and services. In the case of insect outbreaks, we have insects that eat leaves and trees that produce them. A first precondition for understanding dynamic systems is to identify their main constituents.

A second feature common to all of these systems is that interactions among their individual parts do not occur instantaneously, but in a time-delayed manner. Those infected with the flu may keep wandering about for a few days and pass on the virus to others. Producers who offer their goods and services on the market may generate excess supply that leads to a drop in price. As a result, they may restrict production in the next period, leading to shortages and subsequent price increases. As forests grow they provide increasing amounts of food for insect populations that will take on a size that is ultimately too large for the forests to sustain. Thus, a second precondition for understanding dynamic systems is to identify the extent to which interactions among system components are subject to time lags.

A third feature of many real-world dynamic processes is that the response of one system component may not occur in direct proportionality to a stimulus that it receives. Rather, the responses may be related to the square of the initial stimulus or take place in some other nonlinear relationship.

Understanding the world in which we live requires understanding the role of complex feedback processes and the way in which their strength changes over time. Yet modelers are often tempted to compartmentalize systems into subsystems for which it is possible to specify cause–effect relationships that lead to "closed-form" solutions of the kind we are accustomed to seeing in textbooks.

Unfortunately, the methods that achieve such solutions may limit the extent to which one is able to accommodate time lags and nonlinear relationships. By placing undue emphasis on finding closed-form solutions, we run the risk of eliminating from our models the very features that make them interesting.

Fortunately, computer technology provides alternative tools, ones that enable us to put more life into our models of real world processes. In this article I will

Art by Sergey Ivanov

introduce you to a graphical programming language that enables virtually anyone to describe, model, and analyze complex dynamical systems. As such, this software provides a valuable tool for efforts to understand the world in which we live. It's a starting point for the investigation of nonlinear dynamic phenomena and provides an opportunity to easily assess the limits of our knowledge about these phenomena, to foster dialog about them, and to generate new knowledge.

Systems thinking and dynamic modeling

Every day we develop models of dynamic processes. When trying to cross a busy street, we estimate the width of the street, our own speed, and the speed of the cars approaching us. In our mind we abstract away details that we consider inconsequential, such as the color of the cars. Then, we relate the remaining pieces of information to each other and make a projection of the possible outcome if we try to cross the street. If we conclude that it's safe to walk and have drawn the right conclusion, we'll use our model again in similar situations. If we're wrong—but sufficiently lucky—we'll revise our model for next time.

For some decisions, mental models are sufficiently simple and accurate to provide a basis for action. However, the larger the number of system components and the more time lags and nonlinearities there are in the system, the more difficult it is for us to develop adequate mental models for decision making. For large, complex systems, direct experimentation may also be undesirable. For example, it's much safer and less costly to do global climate change experiments on the computer rather than in the real world.

To model and better understand nonlinear dynamic systems requires that we describe the main system components and their interactions. System components can be described by a set of "state variables"—we'll also call them stocks—such as the number of people in a country, the mass of an organism, or the amount of capital in an economy. These state variables are influenced by "flows," such as the births and deaths that occur in a population, growth of an organism, or investment in new capital. The size of the flows may in turn depend on the stocks themselves and other parameters of the system.

There exist various programming languages that are specifically designed to facilitate modeling of nonlinear dynamic systems. Among the most versatile of these languages is the graphical programming language STELLA®. To model the dynamics of a system in STELLA, we begin by identifying the system's stocks, flows, and parameters, and then establish the appropriate connections among them.

STELLA represents stocks, flows, and parameters (also called "converters") with the symbols shown in figure 1. By selecting one of these symbols from the STELLA toolbar and placing it in a diagram window, we specify the components of our model.

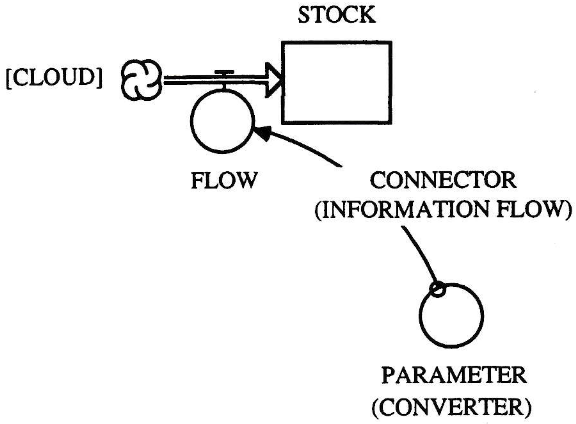


Figure 1

Let's assume we want to model the spread of a virus in a town with an initial population of 1,020. There are two categories of people in this town—a group of 1,000 people that are not immune to a disease and a group of 20 people who carry the virus. These two groups of people are our state variables and constitute the first part of our dynamic model (fig. 2). After assigning names to these stocks, we address the question marks indicating that the size of each stock is not yet specified. Double-clicking on each of the stocks opens a dialog window in which you are asked to specify the stock's initial size. This provides the opportunity to enter population sizes of 1,000 and 20 for the respective stocks. If we now click OK, the stock's dialog box closes and the question mark is removed.

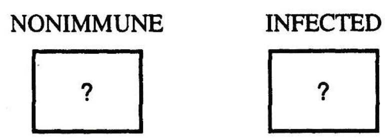


Figure 2

Next we need to specify how these stocks change over time—for example, that there is an influx of nonimmune people each week into the town that we model here. Let's assume that there are seven immigrants into the system To capture this immigration, we select the "flow" symbol from the toolbar, place it in the diagram, and drag the arrow onto the NONIMMUNE stock. The fact that these nonimmune immigrants come from a place that is not part of the model is represented by the fact that the flow originates in a cloud. Next we specify the size of this flow, in our case by entering "7" in the flow symbol's dialog

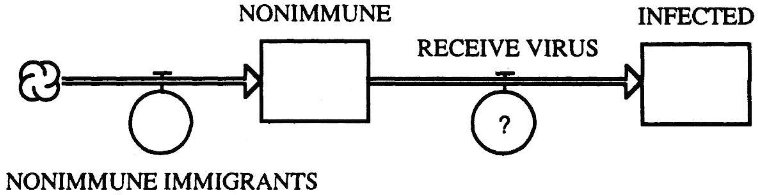


Figure 3

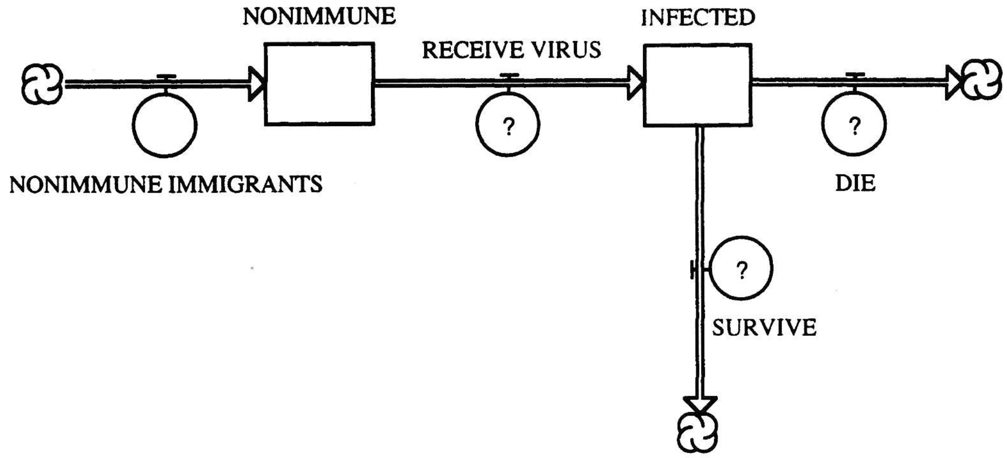


Figure 4

box. Similarly, we include a flow RECEIVE VIRUS to model the rate at which people from the nonimmune stock make the transition to the INFECTED stock of people who carry the virus (fig. 3).

Before dealing with the question mark appearing in the RECEIVE VIRUS flow (we have not yet specified the size of that flow), let's assume that 90% of the people that are infected with the virus survive the disease and that the other 10% die. Those who survive are subsequently immune to the virus and do not infect others. Two more flows are required to move people out of the INFECTED stock of those that carry the virus (fig. 4).

Now we select the STELLA icon representing a parameter, such as the survival rate. By double-clicking on SURVIVAL RATE we open a new dialog box and specify the value of this parameter as 1/10.

Next, we need to tell STELLA how the survival rate affects the size of various flows. To do this, a fourth modeling tool is required. The information arrow conveys the fact that one part of the model has impact on another. Here, we will specify that the SURVIVAL RATE and the INFECTED stock together determine the number of people that survive each week:

$$\text{SURVIVE} = \text{SURVIVAL RATE} * \text{INFECTED} \qquad (1)$$

Since the death rate is (1 – SURVIVAL RATE), these same icons also determine the flow of people that die each week as

$$\text{DIE} = (1 - \text{SURVIVAL RATE}) * \text{INFECTED} \qquad (2)$$

Our model now looks like the one in figure 5. Note that individuals do not move instantaneously through the system. The model is run for discrete time steps. At each time step, the equations that describe the system's dynamics are executed by the computer. For example, individuals enter as NONIMMUNE IMMIGRANTS and become part of the stock of NONIMMUNE, where they temporarily remain. While they are part of the NONIMMUNE stock, they are susceptible to the disease. Some NONIMMUNE individuals will get removed during the next time step into the stock of INFECTED, where they stay for a time step, until they leave the system either as survivors or dead. By making these simulation time steps small, we can approximate continuous time. But the sequence of the model steps that describe the progression from NONIMMUNE to INFECTED captures the essence of the time-delay at which the disease can be passed on to others.

CONTINUED ON PAGE 34

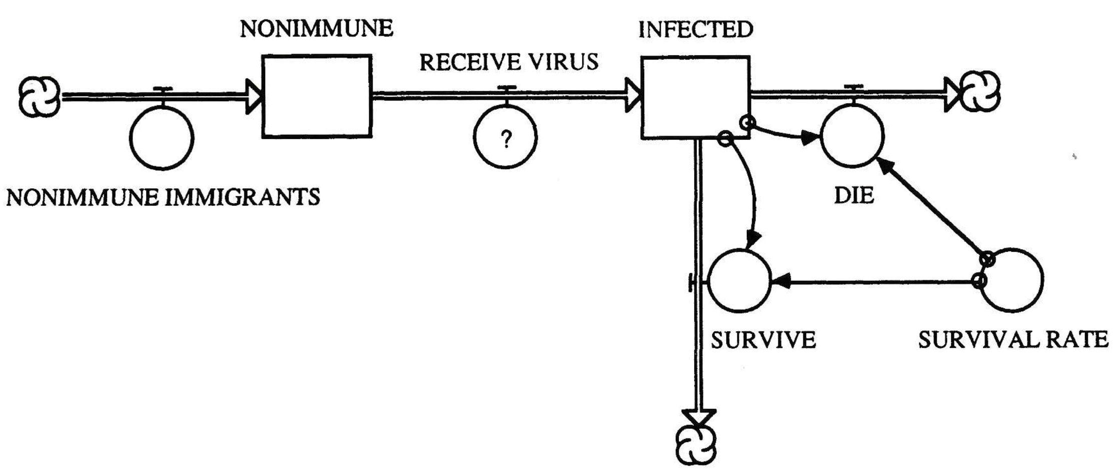
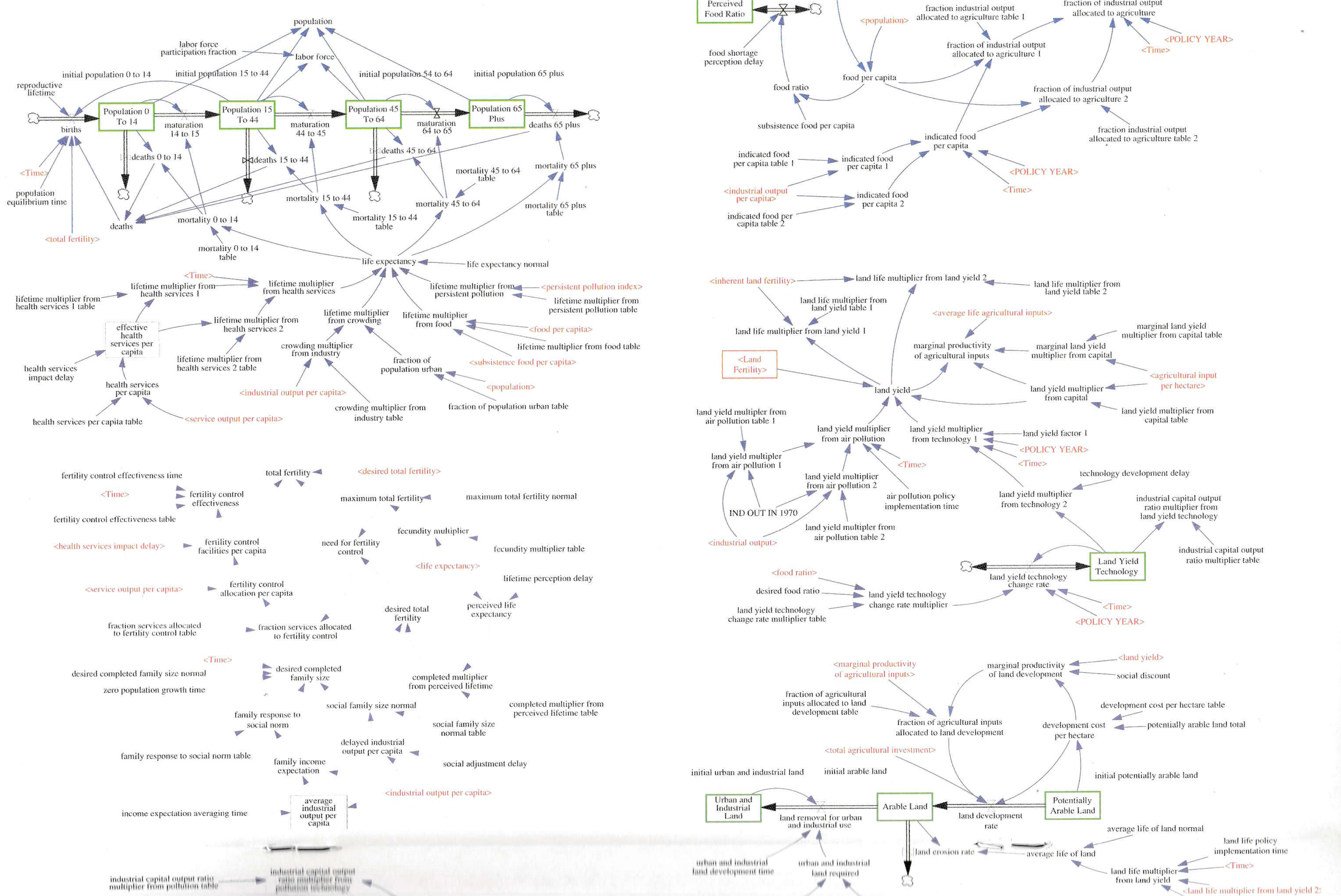
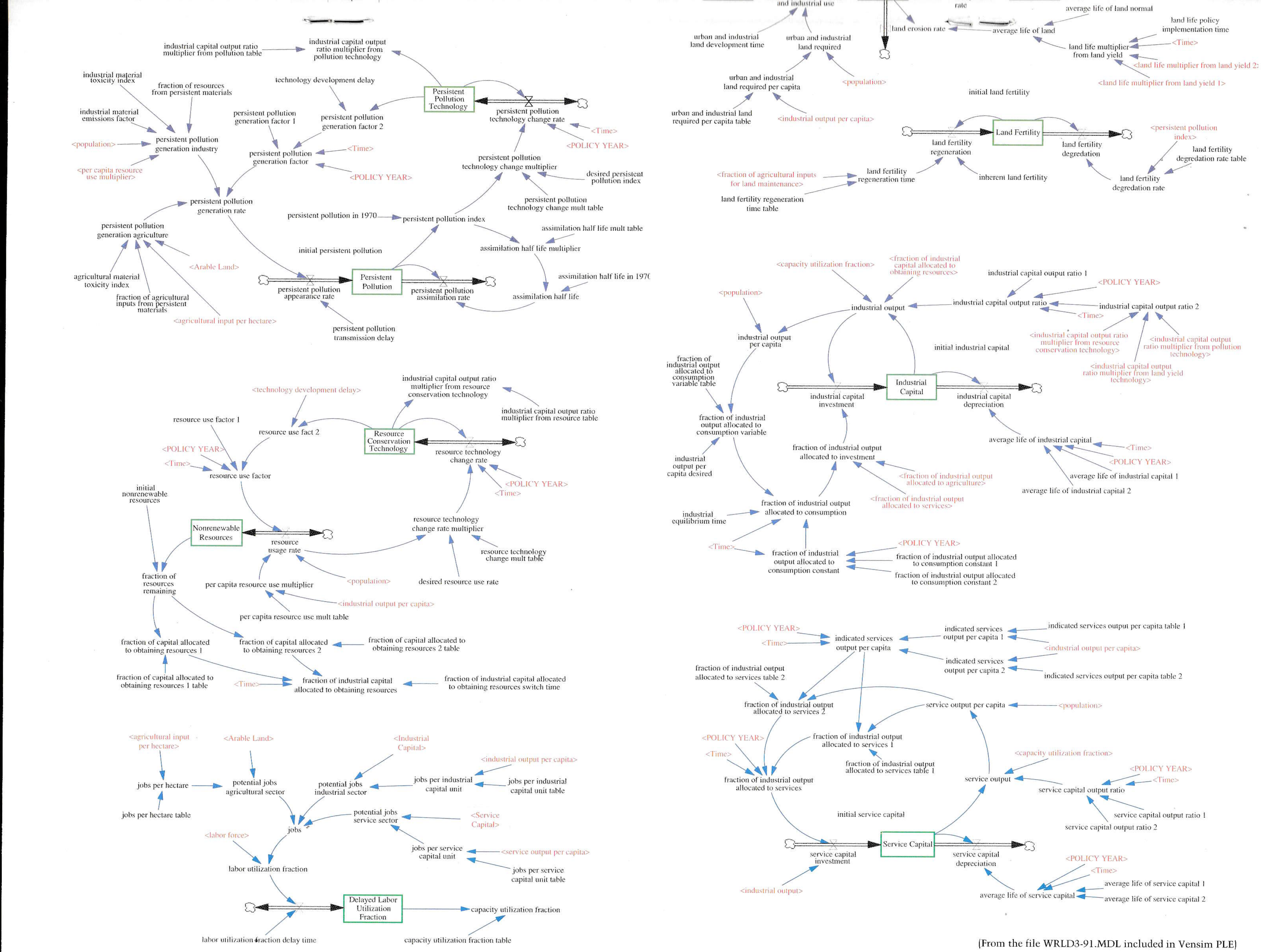


Figure 5

KALEIDOSCOPE

The World3 model





(From the file WRLD3-91.MDL included in Vensim PLE)

"LEARNING FROM A VIRUS" CONTINUED FROM PAGE 31

We still need to specify the mechanism by which the virus gets passed from person to person. Here epidemiologists reason that the number of meetings between nonimmune and infected people is proportional to the size of the populations. A tripling in the size of INFECTED would lead to a tripling in the number of meetings, while a halving would halve the number of meetings. To reflect such a mechanism, we draw information arrows from INFECTED to NONIMMUNE to RECEIVE VIRUS and, in the resulting dialog box, multiply the size of the two populations by a RATE OF CONTACT. This corresponds to

$$\text{RECEIVE VIRUS} = \text{RATE OF CONTACT} * \text{NONIMMUNE} * \text{INFECTED} \quad (3)$$

It remains to define RATE OF CONTACT. Here we'll assume that the rate at which people make contact depends on the size of the stock labeled INFECTED, and this is conveyed by an information arrow from INFECTED to RATE OF CONTACT. We also assume that this rate is high as long as the number of people who do not suffer from the disease is small. However, as more people become sick, the rate of contact declines (the sick people are at home in bed). To build this assumption into our model, we double click on the RATE OF CONTACT icon and enter, in graphical form, a guess of how the contact rate is related to the number of people that carry the virus. The following assumption is built into our model (fig. 6).

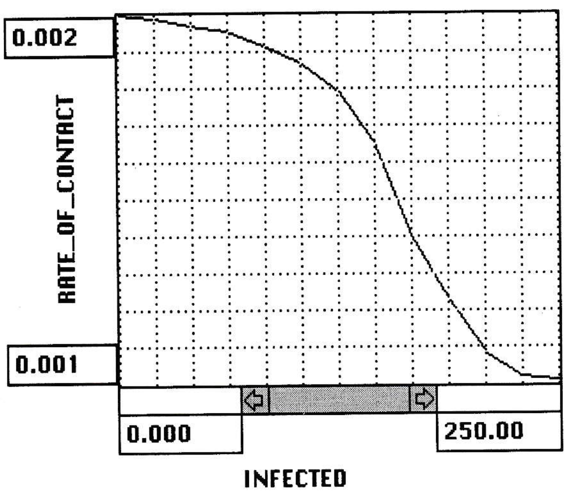


Figure 6

Specifying the above (highly nonlinear) relationship completes the STELLA formulation of the model (fig. 7). The nonlinear relationship between INFECTED and RATE OF CONTACT has led to a model that cannot be solved in closed form. Also, we have no intuitive basis for inferring the course of the disease. However, given software such as STELLA we can obtain a numerical solution for the model we have constructed. Running the above model over 300 simulated weeks yields the results shown in figure 8.

Having obtained a solution, we can summarize the phenomenon as follows. There are periodic severe outbreaks of the disease in which both the stocks of nonimmune people and those that carry the virus grow slowly and steadily. The resulting number of contacts is initially small, but it increases quickly as more nonimmune people enter the town. As the product defined in equation (3) grows, so does the number of people moved from the stocks into the clouds—either they die or they become immune. In both cases, we need not keep track of them because they have no subsequent influence on the spread of the disease. Once these people have left the model, the RECEIVE VIRUS flow gets smaller again, making the disease seemingly disappear, only to reappear some weeks later.

In addition to acquiring an insight into the cyclic nature of the epidemic, we also note that successive peaks are becoming smaller. Can you develop a hypothesis for why, in the very long run, the number of INFECTED stabilizes at a low level? What would that level be? (An answer is provided in the Answers, Hints & Solutions section at the back of the magazine.)

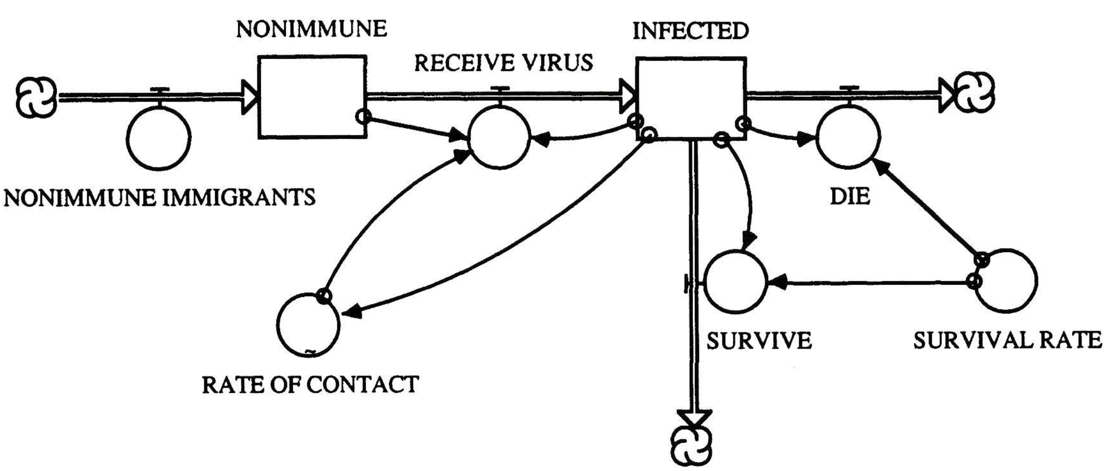


Figure 7

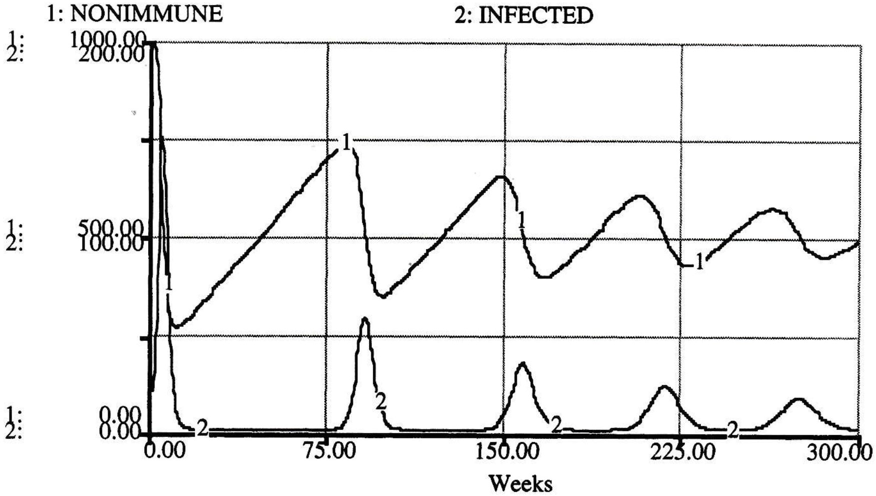


Figure 8

A simple Ebola model

Now that we've gained some insight into one kind of epidemic, we can ask how the outbreaks of the disease would change if the virus were spread from one species to another, similar to the spread of Ebola. Here one population (monkeys *M*) carries a virus that can be spread to a second population (humans *H*). The virus is also passed within the monkey population from infected to nonimmune ones, but it is not received by monkeys from humans. In contrast, humans can get the virus both from monkeys and from other humans that carry the virus.

The STELLA model corresponding to this situation consists of two submodels, each of which is similar to the one already considered. We now have epidemics taking place among monkeys and among humans, with an important additional feature: an "information arrow" from INFECTED M to H RECEIVE VIRUS corresponds to the fact that humans can get the virus from infected monkeys as well as from infected humans. By way of simplification, we'll now assume that the various "rates of contact" are constants rather than depending on the size of the infected populations.

The structure of the resulting STELLA model is shown in figure 9. The fact that monkeys receive the virus only

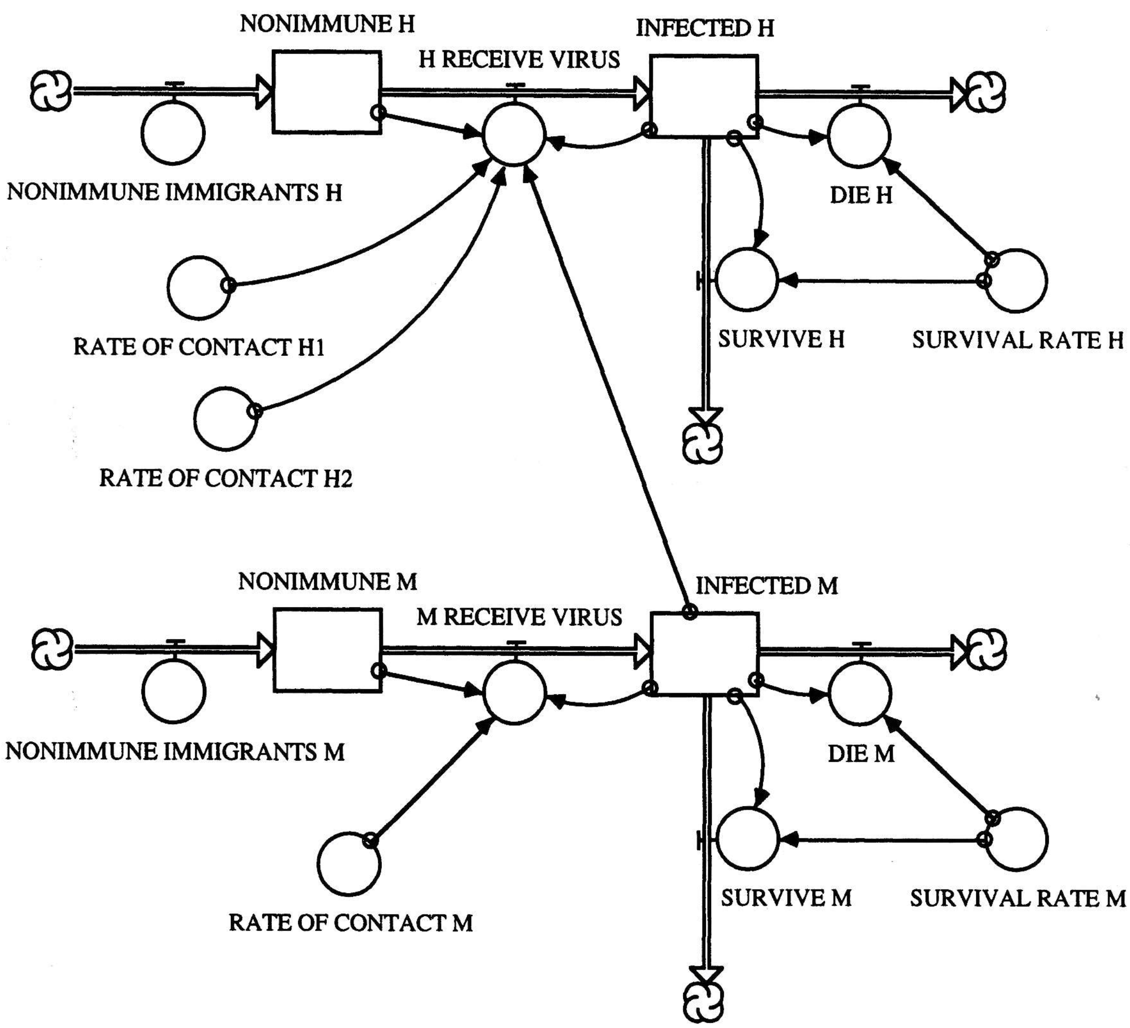


Figure 9

	Humans	Monkeys
NONIMMUNE IMMIGRANTS	7	10
NONIMMUNE	1,000	1,000
INFECTED	5	20
RATE OF CONTACT	H1 = 0.008 H2 varies	0.003
SURVIVAL RATE	varies	0.2

Figure 10

from other monkeys is reflected by the relation

$$\text{M RECEIVE VIRUS} = \text{RATE OF CONTACT M} * \text{INFECTED M} * \text{NONIMMUNE M} \quad (4)$$

The fact that people receive the virus from both monkeys and people is reflected by

$$\text{H RECEIVE VIRUS} = \text{RATE OF CONTACT H1} * \text{NONIMMUNE} * \text{INFECTED H} + \text{RATE OF CONTACT H2} * \text{NONIMMUNE} * \text{INFECTED M} \quad (5)$$

The question marks that originally appeared in figure 9 were removed by specifying the initial size of the four stocks, the various rates of contact, and the survival rates for both monkeys and humans.

Our hypothetical initial conditions and parameters of the model are presented in figure 10. The entry "varies" in figure 10 corresponds to the assumption that there are two different strains of the virus. One strain gets passed from monkeys to humans through direct contact. While the likelihood of this occurrence is low, this is the more deadly strain. The other strain of the virus can be carried through the air from the monkeys to humans. While it has a higher likelihood of affecting humans, this virus is not as deadly as the first.

Such considerations lead us to suggest two model runs based on the following alternative values for RATE OF CONTACT H2 between humans and infected monkeys and the corresponding values for the SURVIVAL RATE H. The first run corresponds to the airborne virus with higher contact rate and higher rate of survival. The second run corresponds to the virus passed by direct contact that has a lower rate of survival (fig. 11). Figure 12 gives the results of these two runs.

Model run	RATE OF CONTACT H2	SURVIVAL RATE H
1	0.00025	0.155
2	0.00015	0.065

Figure 11

The first model run shows that the virus can stay in the human population for long times. The behavior of the disease mirrors that of our very first model above, in which we only considered one population. Periodic outbreaks are followed by long periods in which the disease affects very few individuals. The more intriguing case is depicted by the second model run. Here, a similar pattern of disease outbreak, temporary calm, and new outbreak occurs. But the disease totally disappears after 290 model weeks. Can you explain why? (The answer is given in the solution section.)

As in the previous model, this setup is a very simplistic one, with a very small number of stocks (four) and

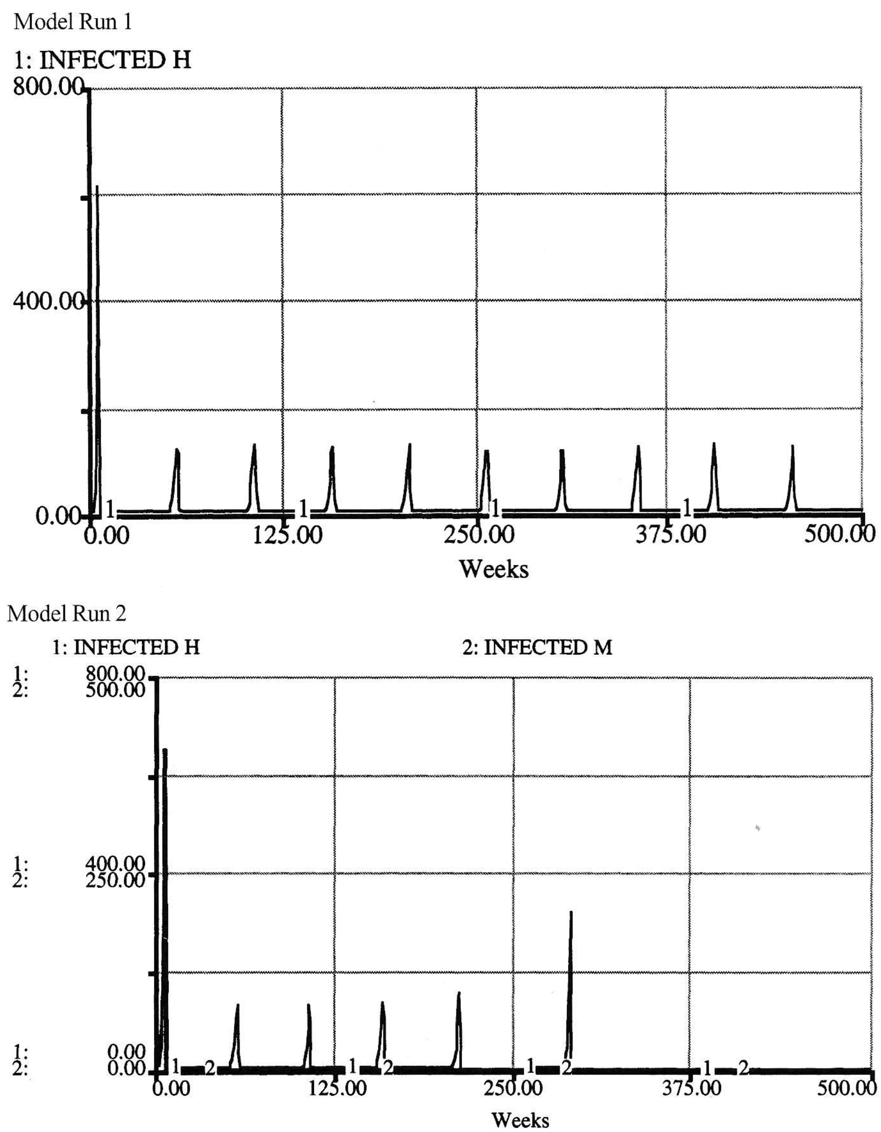


Figure 12

flows (eight) and a very limited number of nonlinearities (two). Yet the dynamics are much richer than many of us would have expected! In one case, the disease shows periodic outbreaks and always a positive number of INFECTED H. In another case, after making small changes in parameters such as the contact rate, we find that the disease entirely vanishes. These findings would suggest that an understanding of contact rates is central to understanding the spread and persistence of a disease in a population over time. These findings also indicate that it may be difficult to respond to outbreaks of highly contagious diseases—there just may not be enough time to find a vaccine, and the disease may disappear soon after wreaking havoc within a population.

The lessons of this model, however, go beyond those findings. By crystallizing our understanding of system processes within the context of a dynamic model, and by putting the pieces of a system together and running them in interaction with each other, we have been able to lay open the dynamic consequences of our assumption and generate insight that would have been very difficult to gain without the help of the model and a computer program like STELLA. The model also helped us identify key parameters of a system's dynamics—such as the contact rate that describes an important aspect of the spread of a disease—and through this may guide data collection and analysis. Along the way, important new questions may have been stimulated whose answers, in turn, could be found with expansions of our model or the development of a new model. ◘

About STELLA®

STELLA is becoming increasingly popular in the social and natural sciences as well as in business decision making. It is used on the Macintosh, IBM, and—in conjunction with compilers—on mainframes and supercomputers. An ever larger number of books is becoming available that provide introductions to systems thinking and dynamic modeling, some of which make extensive use of the STELLA software. It is my hope that through these books and many other collaborative efforts we can build a modeling community of students, teachers, and researchers spreading the dynamic modeling enthusiasm—and systems thinking—by word of mouth and by people in groups of two or three sitting around a computer doing this modeling together, building a new model or reviewing one by another such group.

Matthias Ruth *is a professor at the Center for Energy and Environmental Studies and the Department of Geography, Boston University. He is the author of several books on dynamic modeling published by Springer-Verlag. Dr. Ruth's Web page can be found at http://web.bu.edu/CEES/readmoreMR.html.*

Art by Sergey Ivanov

The far from dismal science

How input–output economics sheds light on environmental issues

by Dean Button, Faye Duchin, and Kurt Kreith

ON ITS MASTHEAD, *QUANTUM* is billed as "The Magazine of Math and Science." Does the "dismal science" of economics have a place in the pages of *Quantum*? In other words, do mathematics and the physical sciences overlap with economics in a meaningful way?

Not unexpectedly, our goal in writing this article is to suggest some affirmative answers to these questions. Furthermore, in light of this issue's focus on computer technology, system dynamics, and *The Limits to Growth*, we will use spreadsheets and STELLA® to make our case in an environmental context.

With STELLA's icons, traditional economic theory can be represented by the diagram in figure 1. Here "economy" corresponds to a very complex system—one that absorbs a wide range of "resources" and transforms them into "goods" that our society consumes. Lately, however, we have become increasingly aware of undesirable by-products that are associated with some economic processes. To reflect this fact, we have labeled the economy's outflow to include "bads" as well as "goods."

More important than the labels in figure 1 is the fact that "resources" originate in a cloud and "goods and bads" terminate in a cloud. These iconic clouds reflect the fact that conventional economic theory tends to sidestep questions of where an economy's resources come from and where its goods and bads eventually go. In other words, much of economic theory is dedicated to studying the economy as if it were a closed system, without considering the larger biophysical world in which the economy exists. In fact, the economy is an open system existing within a much larger, materially closed system. From a traditional perspective, however, STELLA's "clouds" serve as a reminder of the fact that sources and sinks are not included in figure 1.

Against this background, studies such as *The Limits to Growth* can be thought of as posing a fundamental challenge to traditional economics. They ask economists to emulate the physical sciences by embedding their theories within a larger system, notably that of the Earth and its biosphere (see figure 1 of "*The Limits to Growth* Revisited" in this issue).

Actually, efforts to respond to this challenge predate *The Limits to Growth*. Nicholas Georgescu-Roegen began his professional life as a mathematician, one who studied thermodynamics under the tutelage of Emile Borel. After turning his attention to economics, he sought to reconcile the laws of thermodynamics with the functioning of economic systems. In particular, he confronted economic theorists with the Second Law of Thermodynamics, asserting that economic activity (like all other fully contained physical processes) increases the measure of disorder called entropy. Another prominent economist who has responded to

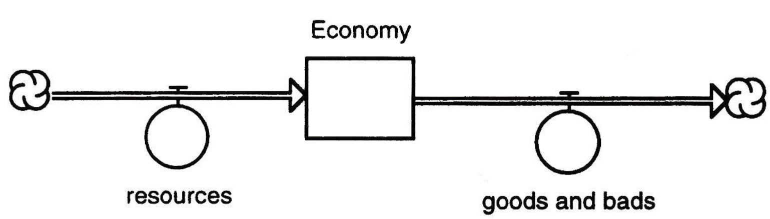


Figure 1

Art by Sergey Ivanov

such challenges is Wassily Leontief. As a Harvard colleague of Georgescu-Roegen, Leontief harnessed the power of computers to develop what he called input–output economics. His earliest efforts were aimed at improving on figure 1 by acknowledging that economies use different categories of resources—termed inputs (iron, coal, petroleum, lumber, labor, etc.) to produce many categories of goods termed outputs (food, machinery clothing, housing, etc.). Here Leontief created mathematical tools to represent the interrelationships among such inputs and outputs. Along the way to winning a Nobel prize for his contributions, he also found time to explain his work in the popular press (in *Scientific American*).

While input–output economics was originally formulated as a closed system, Leontief and his colleagues soon recognized a need to acknowledge the importance of sources and sinks outside of the economic system that is itself situated within the Earth's ecosystem. Input–output economics served as the foundation on which a model of the world economy was developed and subsequently refined. Though the details of the world model are beyond the scope of this article, we will be able to illustrate some of the underlying mathematics.

Input–output economics

In one of his articles, Leontief illustrated input–output analysis in terms of an economy consisting of just three sectors: Agriculture, Manufacturing, and Households. These sectors produce three distinct commodities: wheat (measured in bushels), cloth (measured in yards), and labor (measured in person-years). What input–output analysis provides is a mathematical representation of the interdependence of these sectors. Thus agriculture requires wheat for next year's seeds, cloth for sacks, and labor to till the fields. Manufacturing requires wheat stalks for fiber, cloth to package its products, and labor to work in the mills. Households require wheat for food, cloth for apparel, and labor for various domestic services. The resulting economic model corresponds to figure 1 in that there is no attempt to address where its three categories of "resources" come from or where its three categories of "goods and bads" (in this case there are no bads) eventually go.

To get some insight into the issues arising in a closed economic system, let's begin by replacing the Manufacturing sector in this imaginary economy with an Energy sector. This substitution enables us to retain the simplicity of a three-sector economy while also acknowledging the finiteness of one of its resources (fossil fuels flowing into this economic system) and the fact that CO_2 produced by the burning of fossil fuels must be recycled and/or absorbed by the Earth's ecosystem (fig. 2).

After developing the essential mathetmatical ideas in this three-sector context, we will reintroduce Manufacturing as a fourth sector. At this point it's useful to introduce some additional mathematics—namely, the matrix theory required for the solution of three linear equations in three unknowns. With such tools at our disposal, the transition to four sectors will illustrate the mathematical ideas required for more elaborate applications—for example, the World Model based on forty-four sectors that provides the basis for a recent book, *The Future of the Environment*.

So let's begin with a three-sector economy, one whose components are now labeled Agriculture, Energy, and Households. Assuming an annual output of 100 bushels of wheat, 50 barrels of fuel, and 300 person-hours of labor, we represent the outputs of these three sectors by $x_1 = 100$, $x_2 = 50$, and $x_3 = 300$.

What input–output analysis now calls for is a breakdown as to how these products are distributed among the various sectors. Such information can be conveyed by means of a 3 × 4 grid like table 1. (Note: the form of this table follows that prescribed in matrix algebra, in which matrices are identifed by the *number of rows* and the *number of columns* ($r \times c$) contained in the matrix. Here there are three rows and four columns, represented by the unshaded area.)

Referring to wheat, fuel, and labor as commodities 1, 2, and 3, respec-

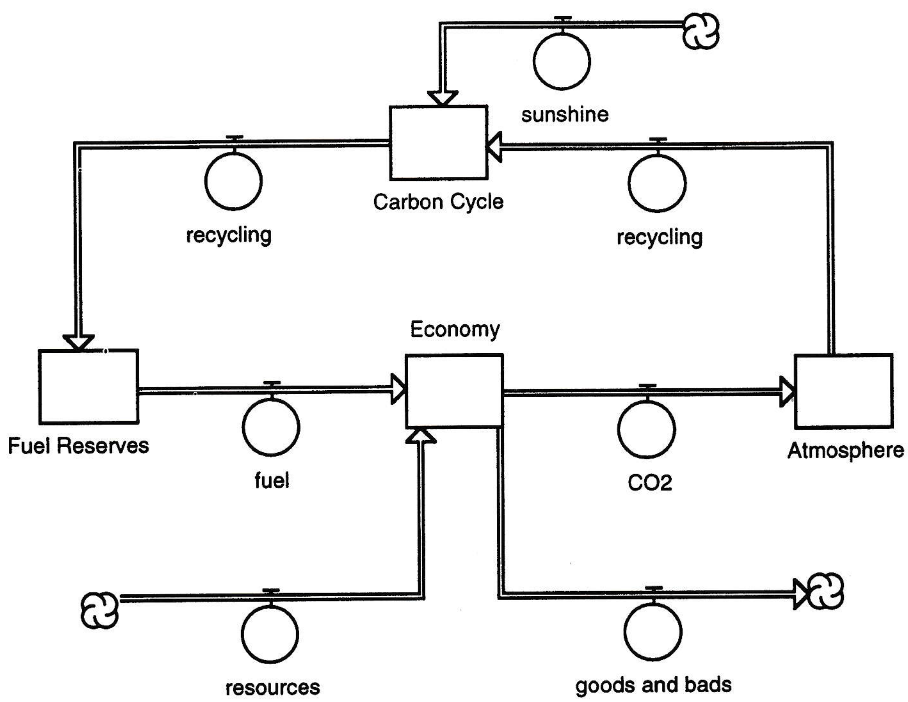


Figure 2

Table 1

	Agricultural requirements	Energy requirements	Household requirements	Totals
Agricultural outputs	x_{11} = 10 bushels	x_{12} = 20 bushels	x_{13} = 70 bushels	100 bushels
Energy outputs	x_{21} = 20 barrels	x_{22} = 10 barrels	x_{23} = 20 barrels	50 barrels
Household outputs	x_{31} = 220 person-hours	x_{32} = 50 person-hours	x_{33} = 30 person-hours	300 person-hours

tively, the table assigns a value x_{ij} to the *number of units of commodity i required to sustain the above output of commodity j*. Here $x_{11} = 10$, $x_{21} = 20$, and $x_{31} = 220$, reflecting the fact that it requires 10 bushels of wheat, 20 barrels of fuel, and 220 person-years of labor to sustain the production of 100 bushels of wheat, and so on.

At this point it becomes important to acknowledge a fundamental difference between the first two sectors (Agriculture and Energy) and the third (Households). Sectors 1 and 2 represent commodities (food and fuel) that are required to sustain the community's well being. According to table 1, this particular community's households require 70 bushels of wheat and 20 barrels of fossil fuel. They are in turn obligated to invest a total of 300 person-hours of labor to sustain this level of consumption. Since $x_{13} = 70$ and $x_{23} = 20$ are *requirements* that a particular level of consumption imposes on an economy, we designate these as *exogenous* (externally imposed) variables.

This special role calls for a change of notation. In what follows, we'll set $x_{13} = y_1$ and $x_{23} = y_2$. As for $x_{31} = 220$, $x_{32} = 50$, $x_{33} = 30$ and $x_3 = 300$, let's assume that the community is somewhat flexible in the amount of labor it's able to commit to sustaining its economy. On this basis, our analysis will for the time being ignore these variables.

These changes enable us to summarize the first two rows in table 1 as

$$x_{11} + x_{12} + y_1 = x_1,$$
$$x_{21} + x_{22} + y_2 = x_2,$$

or

$$(x_1 - x_{11}) - x_{12} = y_1,$$
$$-x_{21} + (x_2 - x_{22}) = y_2. \qquad (1)$$

Note that the exogenous variables $y_1 = 70$ and $y_2 = 20$ have now been isolated on the right side of equations (1).

To simplify this last system of equations further, we make an additional definition of the *structural coefficients*

$$a_{ij} = \frac{x_{ij}}{x_j} \text{ for } i = 1,2 \text{ and } j = 1,2. \qquad (2)$$

Recalling that x_{ij} denotes the amount of commodity *i* required to sustain the production of x_j units of commodity *j*, it follows that a_{ij} denotes the amount of commodity *i* required to sustain the production of a *single unit* of commodity *j*. That is, since it takes 20 barrels of fuel ($x_{21} = 20$) to produce 100 bushels of wheat ($x_1 = 100$), it takes 0.2 barrels of fuel to produce a single bushel of wheat. This is the rationale for introducing the structural coefficient $a_{21} = x_{21}/x_1$. This last change in notation enables us to write the system of equations (1) as

$$(1 - a_{11})x_1 - a_{12}x_2 = y_1,$$
$$-a_{21}x_1 + (1 - a_{22})x_2 = y_2. \qquad (3)$$

Recalling the values of x_{ij} and x_i from table 1, this system becomes

$$0.9x_1 - 0.4x_2 = y_1$$
$$-0.2x_1 + 0.8x_2 = y_2. \qquad (4)$$

Problem 1. Use equations (4) to confirm that an imposition of the exogenous values $y_1 = 70$ and $y_2 = 20$ corresponds to $x_1 = 100$ and $x_2 = 50$.

What makes the system of equations (4) so important is that it embodies the *structure* of this particular economy. If the population of this community grows, the exogenous variables y_1 and y_2 can be expected to grow as well, although not necessarily at the same rate. The question "What effect will such population growth have on the economy?" can then be answered in terms of equations (4) (assuming, of course, that the structure of the economy—that is, the input requirements per unit of output—remains unchanged).

By way of a specific example, suppose that the population of this community were to double. A doubling of the exogenous variables to $y_1 = 140$ and $y_2 = 40$ would require that the entire economy double its annual output, from $x_1 = 100$ to 200 bushels of wheat and from $x_2 = 50$ to 100 barrels of fuel.

But what if this community were to accept energy conservation measures, ones that maintain the households' level of fuel consumption at the current level of $y_2 = 20$ barrels/year—even as the population doubles. If the community's food supply is still to increase to 140 bushels of wheat, such conservation measures would *not* succeed in holding fuel production at the pregrowth level of $x_2 = 50$ barrels/year. Rather, agriculture and energy would both experience increased fuel demands (even though the households do not), and we would now determine x_1 and x_2 by solving the system

$$0.9x_1 - 0.4x_2 = 140,$$
$$-2x_1 + 0.8x_2 = 20. \qquad (5)$$

Problem 2. Solve equations (5) for x_1 and x_2. Explain why x_1 fails to double, even though households require twice as much food. Explain the increase in x_2.

Static vs. dynamic

Readers looking ahead to the task of extending these ideas to four sectors (Leontief studied economies described in terms of hundreds of sectors) may see a need to harness technology to help us solve systems larger than equations (3). But before turning to this task, let's note that the techniques presented so far have been static in nature. Given an economy represented by (4), these equations enable us to determine the outputs x_1 and x_2 that correspond to the exogenous values $y_1 = 70$ and $y_2 = 20$, and later to the exogenous values $y_1 = 140$ and $y_2 = 20$.

For each new set of values for y_1 and y_2, a separate application of equations (4) is called for.

But what if we are interested in determining how this economy will evolve with time? It is here that input–output economics becomes more elaborate, introducing considerations such as the technological changes that take place and the accumulation of buildings, machines, tools, etc., that make it possible to produce a changing mix of goods. But even without confronting such concepts, we can build a dynamic model by formulating specific rules by which y_1 and y_2 are likely to change with time and then asking for the corresponding change in $x_1(t)$ and $x_2(t)$. By way of specific example, let's suppose that household demand for food will grow at 3% per year while household demand for energy will satisfy $y_2(t) = 20 + t$. Assuming the structure of the economy remains constant, how will the total demand for energy change with time?

One way of answering this question is by means of a spreadsheet program that calls for repeated applications of equations (4). Underlying such a program is the fact that the solution of equations (4) is given by

$$x_1 = \frac{0.8y_1 + 0.4y_2}{0.64},$$
$$x_2 = \frac{0.2y_1 + 0.9y_2}{0.64}, \qquad (6)$$

and this enables us to use a spreadsheet to calculate (and plot) the changing values of x_1 and x_2 corresponding to the assumed changes in y_1 and y_2 (fig. 3).

	A	B	C	D	E	F	G
1	Time	y1(t)	y2(t)	x1(t)	x2(t)		
2	1	70	20	100	50		
3	2	72.10	22	103.88	53.47		
4	3	74.26	23	107.20	55.55		
5	4	76.49	24	110.61	57.65		
6	5	78.79	25	114.11	59.78		
7	6	81.15	26	117.69	61.92		
8	7	83.58	27	121.35	64.09		

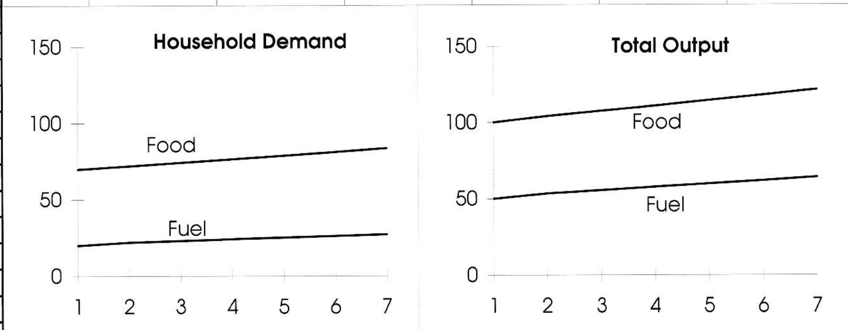


Figure 3

Problem 3. Suppose y_1 grows at 2% per year while y_2 declines at 1% per year. Develop a spreadsheet program that determines the corresponding values of $x_1(t)$ and $x_2(t)$.

Larger systems

Readers familiar with matrices may have observed that equations (6) can be arrived at by other means. That is, equations (4) can be thought of as a matrix equation of the form $\mathbf{Cx} = \mathbf{y}$, where $\mathbf{C}$ is a 2×2 matrix given by

$$\mathbf{C} = \begin{pmatrix} 0.9 & -0.4 \\ -0.2 & 0.8 \end{pmatrix},$$

$\mathbf{x}$ is a column vector with components (x_1, x_2), and $\mathbf{y}$ is a column vector with components (y_1, y_2). In this context, equations (6) now correspond to $\mathbf{x} = \mathbf{C}^{-1}\mathbf{y}$, where

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{C}^{-1}\mathbf{y} = \frac{1}{0.64}\begin{pmatrix} 0.8 & 0.4 \\ 0.2 & 0.9 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

As this example suggests, the problem of solving large systems of linear equations is mainly one of inverting large $n \times n$ matrices.

In order to relate these ideas to the system of equations (3), it's useful to introduce the matrix $\mathbf{I}$ whose elements are all zeros—except for 1's down the principal diagonal. Known as an identity matrix, it enables us to write equations (3) in the form $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{y}$, where $\mathbf{A}$ is the structural matrix whose elements are the coefficients a_{ij} defined by equation (2). In this notation, the solution of equations (3) is $\mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{y}$, where we are now faced with the challenge of computing the inverse of $\mathbf{C} = (\mathbf{I} - \mathbf{A})$.

It's important to note that the information embodied in $\mathbf{I} - \mathbf{A}$ is regularly collected by statistical offices and governmental agencies the world over and is regularly utilized in similar, though more elaborate, economic analyses.

A four-sector economy

With this background in mind, let's reintroduce the Manufacturing sector into the three-sector economy we've been examining. An input–output model for what is now a four-sector economy will confront us with three equations with three unknowns. While there are many ways of solving such systems, our approach will be based on matrix methods that carry over to larger systems as well.

Let's begin by representing our four-sector economy by means of a 4×5 table (see table 2). Noting that $x_{11} = 10$, $x_{12} = 10$, $x_{13} = 20$, ..., while $x_1 = 100$, $x_2 = 40$, ..., our previous method of analysis leads to the system $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{y}$, where

$$\mathbf{I} - \mathbf{A} = \begin{pmatrix} 0.9091 & -0.2 & -0.286 \\ -0.045 & 0.8 & -0.214 \\ -0.182 & -0.4 & 0.8571 \end{pmatrix}$$

and

$$\mathbf{y} = \begin{pmatrix} 70 \\ 20 \\ 20 \end{pmatrix}.$$

Table 2

	Agricultural requirements	Manufacturing requirements	Energy requirements	Household requirements	Totals
Agricultural outputs	10 bushels	10 bushels	20 bushels	70 bushels	110 bushels
Manufacturing outputs	5 yards	10 yards	15 yards	20 yards	50 yards
Energy outputs	20 barrels	20 barrels	10 barrels	20 barrels	70 barrels
Household outputs	220 person-hours	80 person-hours	50 person-hours	30 person-hours	380 person-hours

Readers familiar with matrix multiplication are invited to confirm that

$$(\mathbf{I}-\mathbf{A})^{-1}=\begin{pmatrix}1.2419 & 0.5914 & 0.5618\\ 0.1616 & 1.5054 & 0.4301\\ 0.3387 & 0.828 & 1.4866\end{pmatrix}$$

by showing that $(\mathbf{I}-\mathbf{A})(\mathbf{I}-\mathbf{A})^{-1}=\mathbf{I}$. This fact makes it possible to use $\mathbf{x}=(\mathbf{I}-\mathbf{A})^{-1}\mathbf{y}$ to show that the exogenous values $y_1 = 60$, $y_2 = 20$, and $y_3 = 20$ lead to $x_1 = 110$, $x_2 = 50$, and $x_3 = 70$.

Using a spreadsheet, we can again create a dynamic version of this four-sector model. The following program assumes that y_1 grows at 3% per year, y_2 grows at 1% per year, and y_3 remains constant at 20 barrels/year (see figure 4).

Our common future

There is a growing consensus within the scientific community that in many (but not all) cases traditional, single-disciplinary approaches are not effective in addressing the problems associated with a complex and highly interconnected world. More and more, we've come to appreciate that our world comprises many interrelated and overlapping systems, and that to isolate one or two pieces from the rest of the puzzle may work "in theory" but not in the real world. Instead, it's much more fruitful—though clearly more of a challenge—to look at issues from a variety of points of view and attempt to solve them in the context in which they exist. For some, this broad integrative approach calls for a radical shift in perspective.

Recall how radically different the Copernican notion of planets revolving around a stationary Sun was from the conventional wisdom of the day. Much like the framework that helped guide those after Copernicus, *The Limits to Growth* has helped many to see how important it is to confront many of today's problems from a "systems perspective."

Input–output economics is a powerful tool for taking a more systemic approach to issues involving the use of natural resources, their transformation into goods that human societies need and want, and

	A	B	C	D	E	F	G	H	I	J
1		Food Inputs	Manuf. Inputs	Fuel Inputs	Labor Inputs	Total				
2	Food Output	10	10	20	70	110		0.0909	0.2	0.2857
3	Manuf. Output	5	10	15	20	50	A=	0.0455	0.2	0.2143
4	Fuel Output	20	20	10	20	70		0.1818	0.4	0.1429
5										
6		0.9091	-0.2	-0.286				1.2419	0.5914	0.5618
7	(I-A)=	-0.045	0.8	-0.214	(I-A) Inverse =			0.1613	1.5054	0.4301
8		-0.182	-0.4	0.8571				0.3387	0.828	1.4866
9										
10	Det(I-A)=		0.4831							
11										
12	How will changing household requirements affect total outputs?									
13	Time	1	2	3	4	5	6	7		
14	y1(t)	70	72.1	74.263	76.491	78.786	81.149	83.584		
15	y2(t)	20	20.2	20.402	20.606	20.812	21.02	21.23		
16	y3(t)	20	20	20	20	20	20	20		
17										
18	x1(t)	110	112.73	115.53	118.42	121.39	124.45	127.6		
19	x2(t)	50	50.64	51.293	51.959	52.64	53.334	54.043		
20	x3(t)	70	70.877	71.777	72.7	73.648	74.621	75.62		
21										

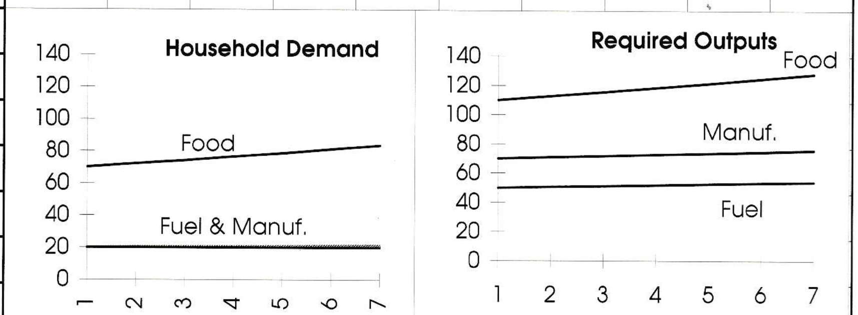


Figure 4

how the ways we provide for ourselves materially impact the rest of the world.

In this article we've seen how input–output economics allows us to construct a mathematical model of an economic system in terms of different sectors, analyze the interrelationship between inputs and outputs in each of the sectors, and provide a formal mechanism for considering where resources for production processes come from and where the products made and their associated wastes eventually go. This helps economics achieve the important objective we mentioned earlier—that of embedding economic theory in a larger, material system. In addition, input–output economics helps both natural and social scientists draw a more sharply focused and detailed picture of what actually takes place in the real world. The urgency of many of today's problems demands solutions of much greater specificity than traditional methods can provide.

Ten years ago, a report entitled *Our Common Future*—more commonly referred to as the Brundtland Report—was prepared for the World Commission on Environment and Development. The Brundtland Report was largely responsible for popularizing the term *sustainable development*, which it defines as the ability of humanity "to ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs." The report describes the environmental and economic problems the world community faces, identifies a number of technological and organizational measures that might be implemented to achieve sustainability, and concludes that two seemingly contradictory goals—economic growth and environmental preservation—can, in fact, be achieved through the appropriate management of technology and social organization. When these goals are considered within the context of ongoing efforts to raise the material standard of living for a growing global population, it becomes obvious that closer scrutiny is warranted in order to learn whether or not achieving them is feasible.

Beyond Brundtland

In *The Future of the Environment* (Oxford University Press, 1994), Faye Duchin and Glenn-Marie Lange use input–output economics to take a closer look at some of the recommendations of the Brundtland Report. Their economic analysis reveals that many of the positions taken are overoptimistic or unrealistic. Duchin and Lange go on to apply input–output economics in developing and evaluating a number of options—called scenarios—that could serve as alternative paths to achieving sustainable development. While traditional economic analysis tends to focus its attention on revealing the "right prices" that would lead to the most efficient allocation of resources in order to satisfy consumer demand and preferences, this analysis seeks to accomplish more. The approach demonstrated in *The Future of the Environment* attempts to go beyond these narrow considerations to integrate a broader systems perspective and provide the empirical evidence needed to make informed and realistic decisions about important issues.

Duchin and Lange begin by constructing several development scenarios: the Our Common Future scenario—based on the assumptions and recommendations set forth in the Brundtland Report; a Reference scenario—based on the assumption that no technical changes occur after 1990 to improve environmental conditions; and three additional scenarios, all based on many of the assumptions contained in the Our Common Future scenario but each assuming a more prominent role for alternative energy sources (either hydroelectric, nuclear, or solar) considered in combination with a more rapid rate of modernization in energy-intensive sectors of the largest developing economies in the world (China and India). By providing methods and data to actually evaluate each of these different development paths, this kind of work extends beyond an exclusive focus on prices and in so doing presents a more realistic picture of what actually takes place in the real world economy.

Economists (like mathematicians and physicists) must make assumptions. But, as *Quantum* readers have seen many times, making assumptions can be a risky business. As in other disciplines, it's important to explicitly recognize the assumptions made because they can—and usually do—play a powerful role in determining the conclusions that are drawn. For instance, the Brundtland Report assumes that clean and efficient modern technologies associated with the use of energy and materials in production processes are adopted in all parts of the world economy over the next several decades. Duchin and Lange assembled the input–output tables and other data necessary to analyze the impact of this assumption and found that this assumption was too optimistic. They looked specifically at the emission levels of three pollutants associated with the production of energy: carbon, sulfur, and nitrogen. They discovered that while levels of emissions in their models are reduced significantly below what they would have been, assuming no technological changes or improvements in production processes (an assumption of the Reference Scenario that Duchin and Lange constructed to serve as a basis for comparison), these emissions still increase by a substantial amount. For instance, carbon emissions increase by a significant 60%. Nitrogen emissions rise by 63%, and sulfur emissions increase 16%.

A sobering scenario

The analysis performed by Duchin and Lange provides an even more detailed look at the future presented in the Brundtland Report. The picture that emerges is not only quite different, it's more structured and systematic. Their data show that, in addition to an overall rise in the level of pollution worldwide, the primary source of pollution shifts from the developed northern hemisphere to the rapidly developing southern hemisphere.

While it may be a comfort to some that pollution levels decline in the north, from a systems perspective it's obvious that the entire world feels the impact of pollution. This isolated "fact" provides little basis for long-term optimism when considered in a broader context. In contrast to the optimism expressed in the Brundtland Report, a significant amount of empirical evidence—including that presented here—indicates that the goal of cleaner air and water will demand far greater efforts to achieve than one might conclude from the Brundtland Report.

With the growing world population, declining stocks of natural resources, environmental degradation, habitat destruction, and climate change becoming regular features in our daily news, there has never been a more urgent need for ways in which we—as individuals and communities—can make better, more intelligent decisions about the direction our future and the future of the environment will take in the opening decades of the 21st century. There are many exciting opportunities to become actively engaged in developing the kinds of innovative and effective tools described here—tools that can make a positive contribution to solving today's environmental and economic predicament. The need to construct new scenarios that help define alternatives to the options currently offered is virtually boundless. By combining the rigorous quantitative analysis provided by input–output economics with the creativity and imagination of a new generation of researchers, there is reason to be hopeful that realistic and feasible answers can indeed be found.

Economics—the dismal science? Far from it! ◘

Dean Button *is Director of Program Development for the School of Humanities and Social Sciences at Rensselaer Polytechnic Institute.* **Faye Duchin** *is the Dean of Humanities and Social Sciences at the same institution and is the author of* The Future of the Environment. **Kurt Kreith** *is a professor at the University of California–Davis and the editor of the LTG articles in this special issue of* Quantum.

PHYSICS CONTEST

Cool vibrations

"The world is never quiet, even its silence eternally resounds with the same notes, in vibrations which escape our ears."
—Albert Camus

by Arthur Eisenkraft and Larry D. Kirkpatrick

A LOW RUMBLE THROUGH the Earth convulses a highway like a fish gasping for air. A child in a distant playground gracefully moves his body, propelling the swing to new heights. An operatic singer shatters a crystal glass with the precision of her voice.

The Earth, the child, and the soprano play with oscillations. Not content with the simple vibrations of sound, or a mass on a string, or a screen door swinging to and fro, this active cast of characters forces the systems and produces fascinating results. To understand what is happening, we will review the simplest vibrating system before exploring the more complex activities of our players.

A mass hangs from a massless spring. In its stable position, the force of gravity on the mass must be equal and opposite to the force of the spring. This defines the equilibrium position of the system x_0:

$$mg - kx_0 = ma = 0,$$

$$x_0 = \frac{mg}{k}.$$

If the system is pulled below its equilibrium position, there is a net force pulling the mass upward. We can define the stretch of the spring x as the distance beyond the equilibrium position:

$$-kx = ma = m\frac{d^2x}{dt^2}.$$

We can "guess" at the solution to this differential equation:

$$x = A\cos(\omega t + \phi).$$

Taking derivatives, we find

$$v = \frac{dx}{dt} = -A\omega\sin(\omega t + \phi),$$

$$a = \frac{d^2x}{dt^2} = -A\omega^2\cos(\omega t + \phi).$$

To find out if we have guessed successfully, we substitute these solutions for x and d^2x/dt^2 into our original equation $-kx = md^2x/dt^2$ and find that this is indeed a solution when

$$\omega = \sqrt{\frac{k}{m}},$$

where $\omega/2\pi$ is equal to the frequency of vibration ν of the oscillator.

The situation gets more complicated when it is made more realistic. All oscillating systems have a retarding force. Often this retarding or damping force is proportional to the velocity of the mass. The amplitude of this oscillation will decrease and decrease until the mass eventually comes to rest. But does the frequency change under this force? The equation of motion is now

$$-kx - b\frac{dx}{dt} = m\frac{d^2x}{dt^2}.$$

A solution to this equation if b (the coefficient of the damping force) is small is

$$x = Ae^{-bt/2m}\cos(\omega' t + \phi),$$

$$\omega' = 2\pi\nu' = \sqrt{\frac{k}{m} - \left(\frac{b}{2m}\right)^2}.$$

A plot of this equation shows an oscillation of constant frequency, which damps out exponentially. Exploring this solution will be part of the contest problem.

A more interesting motion occurs when a varying external force also drives the oscillating mass with a damping force. This external force can be the positioning of a child's body to make the swing reach new heights or the soprano's voice driving the molecules in the glass crystal. In this case, the equation of motion is

$$-kx - b\frac{dx}{dt} + F_m\cos(\omega'' t) = m\frac{d^2t}{dt^2}.$$

The solution of this equation is

Art by Tomas Bunk


TOM BUNK

$$x = \frac{F_m}{G}\sin(\omega'' t - \phi),$$

where

$$G = \sqrt{m^2\left(\omega''^2 - \omega^2\right)^2 + b^2\omega''^2}$$

and

$$\phi = \cos^{-1}\frac{b\omega''}{G}.$$

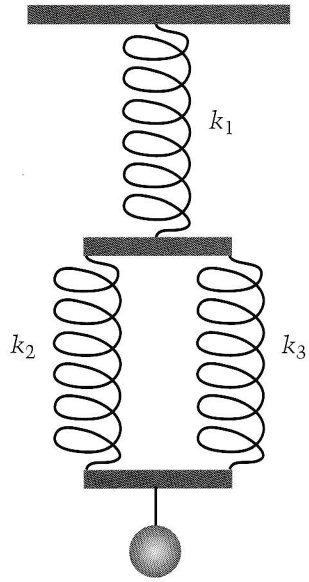
We can see that the frequency of oscillation is now that of the external driving frequency and not the natural frequency. If the driving frequency is equal to the natural frequency and the damping force is zero ($b = 0$), we see that G becomes zero and the displacement x will get infinitely large. This is called resonance. Of course, there is never a situation where the damping force is exactly zero, and we find that the resonance does produce very large displacements, although not infinite. This is the explanation for the collapsing highway, the crystal being broken, and the child being able to swing to such new heights.

Our contest problem oscillates from some simple problems to some graphical and mathematical analysis.

The first problem was a small part of one of the questions in this summer's very successful International Physics Olympiad in Sudbury, Canada. In fact, the local radio station offered a prize to any listeners who could call in a solution.

A. A mass hangs from a massless spring and oscillates with a frequency of 1 Hz. If the spring is cut in half, what is the new oscillation frequency?

B. A mass m hangs from 3 massless springs as shown in the figure. The springs have spring constants k_1, k_2, and k_3. When m is displaced from its equilibrium position, what is the period of oscillation?



C. (1) Sketch the solution for the damped oscillator. (2) The mean lifetime is defined as the time it takes for the oscillator's amplitude to reach $1/e$ of its initial value. Derive an equation for the mean lifetime. (3) A hanging block with a mass of 2 kg is attached to a massless spring with spring constant equal to 10 N/m. The mass is displaced from its equilibrium position by 12 cm. If the damping force has a b value of 0.18 kg/s, find the number of oscillations made by the block during the time interval in which the amplitude falls to 1/4 of its original value. (4) Derive an expression for the velocity of the mass at any given time.

D. A forced oscillator can have substantially different effects on a mass. (1) Show graphically how the amplitude depends on the ratio of the driving frequency ω'' and the natural frequency ω for the following values of the damping coefficient b: $b = 0$, $m\omega/4$, $m\omega/2$, $m\omega$, and $2m\omega$. (2) Derive an expression for the velocity of the mass at any given time.

Please send your solutions to *Quantum*, 1840 Wilson Boulevard, Arlington VA 22201-3000 within a month of receipt of this issue. The best solutions will be noted in this space.

Mars or bust!

As predicted, the Mars Pathfinder landed on Mars on July 4, 1997, and has sent back fantastic pictures of the surface of Mars. We fully expect to find Sojourner toys in sandboxes and backyards very soon.

Our contest problem in the March/April issue explored two simplified ways of leaving Mars orbit and landing on the Martian surface. Excellent solutions were submitted by André Cury Maiali and Gualter José Biscuola from Brazil.

A. The gravitational force acting on a satellite of mass m orbiting Mars is given by

$$F = \frac{GMm}{(R+h)^2},$$

where G is the gravitational constant, M and R are the mass and radius of Mars, and h is the altitude of the orbit above the surface. Because the surface gravity on Mars is given by

$$g = \frac{GM}{R^2},$$

this force can also be written as

$$F = \frac{mgR^2}{(R+h)^2}.$$

Equating either of these expressions for the gravitational force to the centripetal force $mv_0^2/(R+h)$ allows us to solve for the orbital velocity v_0:

$$v_0 = R\sqrt{\frac{g}{R+h}} = \sqrt{\frac{GM}{R+h}}.$$

Using the numerical values given in the problem, $v_0 = 3.38$ km/s.

B. According to the statement describing the first method of landing, the path is an ellipse that is tangent to the orbit and to the Martian surface at the two ends of the ellipse. Let the speed of the satellite be v_X after the retrorockets have fired and v_A at the surface and write down the expressions for the conservation of angular momentum

$$mv_A R = mv_X(R+h)$$

and the conservation of mechanical energy

$$\frac{1}{2}mv_A^2 - \frac{GMm}{R} = \frac{1}{2}mv_X^2 - \frac{GMm}{R+h}.$$

Solving for v_X we obtain

$$v_X = v_0\sqrt{\frac{2R}{2R+h}} = 3.29\ \text{km/s}.$$

Conservation of angular momentum now tells us that

$$v_A = v_X\frac{R+h}{R} = 3.65\ \text{km/s}.$$

This answer makes sense, because the speed must increase as the satellite descends to the surface.

C. This time the rockets fire in the radial direction. Therefore, the rockets do not change the angular momentum of the satellite. Conservation of angular momentum tells us that

$$mv_BR = mv_0(R+h).$$

Thus

$$v_B = v_0\frac{R+h}{R} = 3.74\ \text{km/s}.$$

We can now use conservation of mechanical energy to find the speed at point X. Let's call this speed v_Y to distinguish it from the speed we obtained for the first method. Then

$$\frac{1}{2}mv_Y^2 - \frac{GMm}{R+h} = \frac{1}{2}mv_B^2 - \frac{GMm}{R}.$$

Solving for v_Y and using our expression for v_B, we obtain

$$v_Y = v_0\sqrt{1+\frac{h^2}{R^2}} = 3.40\ \text{km/s}.$$

D. Our Brazilian readers point out that we can get much better comparisons by calculating the changes in the velocities at point X algebraically rather than using our numerical results. In the first case, v_0 and v_X lie along the same direction and we can simply subtract them to obtain

$$\Delta v_A = v_0 - v_X = v_0\left(1-\sqrt{\frac{2R}{2R+h}}\right)$$

$$= 87.7\ \text{m/s}.$$

In the second case, the two velocity vectors $\mathbf{v}_0$ and $\mathbf{v}_Y$ are not parallel, but we know that $\mathbf{v}_Y = \mathbf{v}_0 + \mathbf{v}_r$, where $\mathbf{v}_r$ is the radial component of velocity imparted by the rocket engines. Therefore, we can use the Pythagorean theorem to find

$$\Delta v_B = v_r = v_0\frac{h}{R} = 365\ \text{m/s}.$$

Then $\Delta v_A/\Delta v_B = 0.240$, or about one-fourth as much.

E. We now compare the landing speeds:

$$v_A = v_X\frac{R+h}{R} = v_0\left(\frac{R+h}{R}\right)\sqrt{\frac{2R}{2R+h}},$$

$$v_B = v_0\left(\frac{R+h}{R}\right).$$

Thus

$$\frac{v_A}{v_B} = \sqrt{\frac{2R}{2R+h}} = 0.974.$$

AT THE BLACKBOARD

An ant on a tin can

Finding the shortest path from A to B

by Igor Akulich

TWO STUDENTS ARE PONdering the following problem: A tin can takes the form of a right cylinder with radius R and height H. An ant is sitting on the border circle of one of its bases (point A in figure 1). It wants to crawl to the most distant point B at the border circle of the other base (symmetric to A with respect to the center of the tin). Find the shortest path for the ant.

"But it's a very simple problem!" the first student says confidently. "We just have to consider the planar development of the tin. Let's say, for the sake of definiteness, that the ant first crawls along the side surface and then across the upper base (of course, it's possible that the ant takes the symmetric route: first crawling across the lower base, then along the side surface; but the length of this route is the same). Developing the tin can on the plane (fig. 2a), we see at once that the shortest path goes first along the linear element AM of the cylinder, then continues along the diameter MB. The length of this route is $S_{min} = H + 2R$."

"Wait a second," the other student replies. "One can very comfortably develop the tin can in a different way! Just throw away the lids and spread the side surface on the plane so that we get a rectangle (fig. 2b). Then the shortest path will be the segment connecting points A and B—its length is $S_{min} = \sqrt{H^2 + \pi^2 R^2}$. And the image of the ant's route on the tin can will be a part of a corkscrew line."

The students were on the verge of quarreling when a the idea occurred to them to compare both paths—the shorter one will be the correct answer. First they determined the conditions whereby the lengths of both routes are equal, writing down the equality

$$H + 2R = \sqrt{H^2 + \pi^2 R^2}.$$

Then they transformed it:

$$(H + 2R)^2 = H^2 + \pi^2 R^2,$$
$$H^2 + 4HR + 4R^2 = H^2 + \pi^2 R^2,$$
$$4H = (\pi^2 - 4)R,$$

finally arriving at

$$\frac{H}{R} = \frac{\pi^2}{4} - 1 \cong 1.467.$$

Thus the lengths are equal when

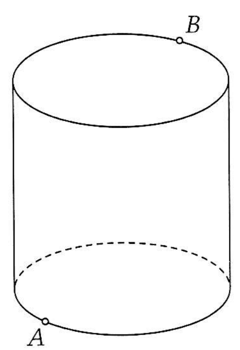


Figure 1

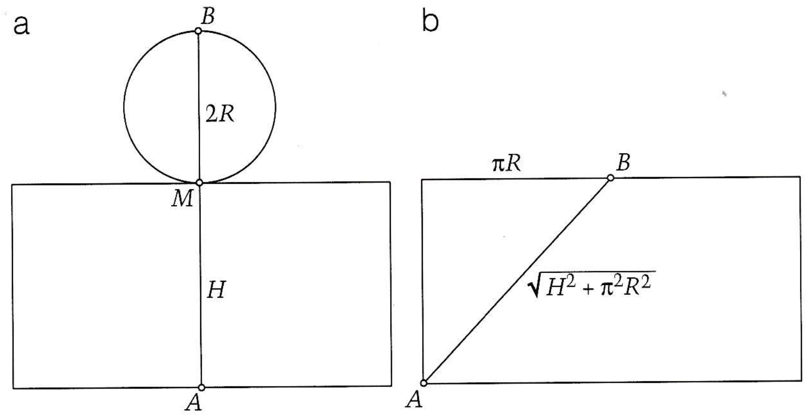
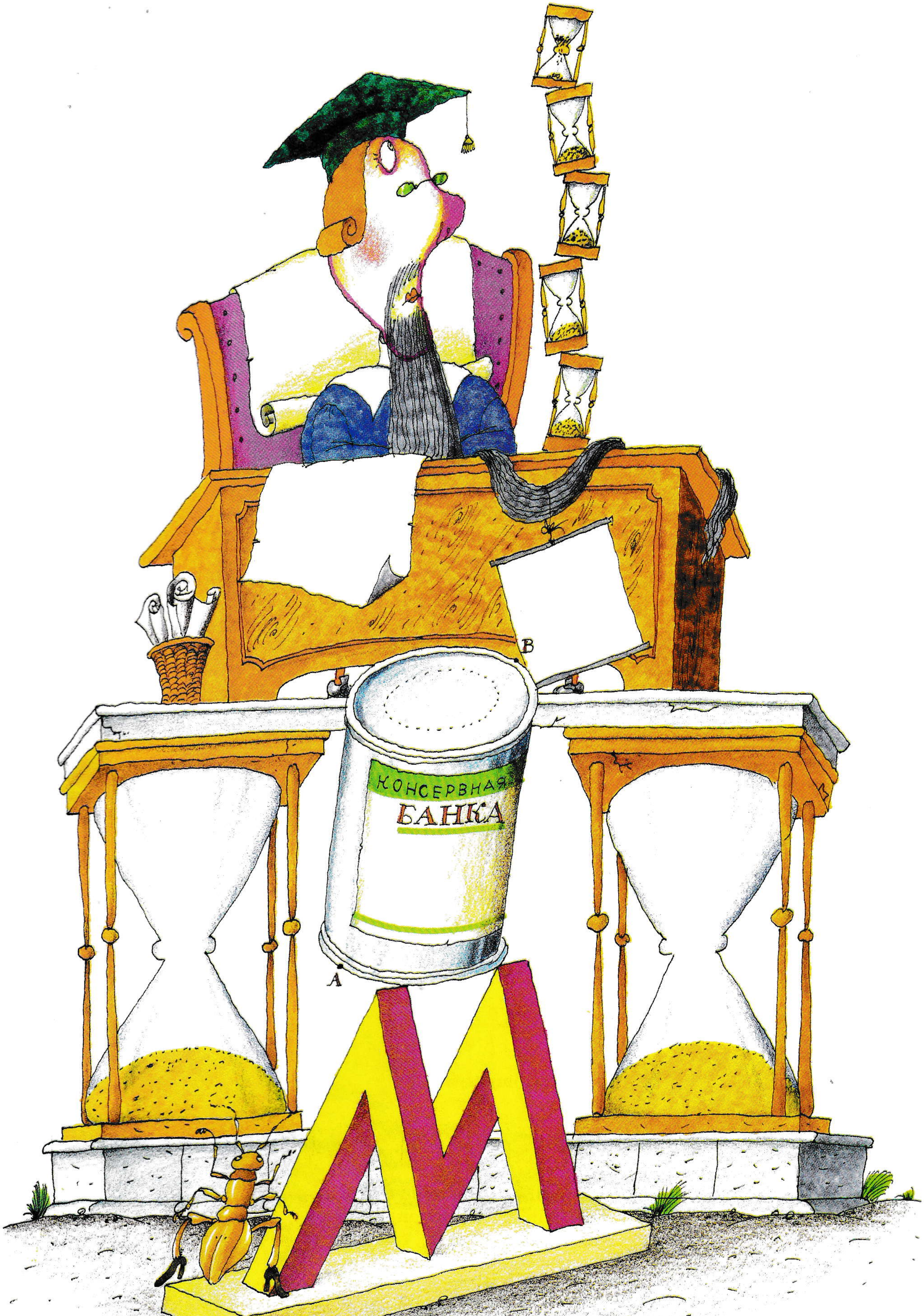


Figure 2

Art by V. Ivanyuk


КОНСЕРВНАЯ
БАНКА
В
А

the ratio of the tin can's height to its radius is equal to this figure. Now we can conclude that if $H/R < \pi^2/4 - 1$, then the shortest path was given by the first student; if $H/R > \pi^2/4 - 1$, then the shortest path was given by the second student.

The students were so glad and proud of themselves, they couldn't help boasting to their math teacher about the solution they'd found.

Question to the reader: What would you tell them if you were their teacher?

The teacher speaks

Let's answer together: "Sorry, you're wrong! You've certainly solved a problem, but . . . it's another problem. Namely, you took two possible routes from point A to point B and determined the conditions whereby one of them is shorter than the other. But in addition to these two, there are many other paths (see figure 3) that go from A along the side surface to an arbitrary point P on the border circle of the upper base, and then from P to B along a segment of a straight line on the upper base."

As you can see, the paths proposed by the students are only particular cases of the path we've suggested. In the path proposed by the first student, P coincides with point M; in that of the second, it coincides with B. That is to say, the students suggested two "extremal" variants. And the truth is, undoubtedly, somewhere in between.

Let's work this through

We're looking for the shortest path. Denoting the center of the upper base by O and the radian measure of the angle MOP by ϕ, we find without much trouble that the length of arc MP equals $R\phi$, and the length of the shortest curve AP is $\sqrt{H^2 + R^2\phi^2}$ (according to the Pythagorean theorem), and the length of segment PB is $2R\cos(\phi/2)$. Thus the length of the entire path is a function of ϕ—namely

$$S = \sqrt{H^2 + R^2\phi^2} + 2R\cos\frac{\phi}{2}.$$

Now we just have to find the minimum of this function at the segment $\phi \in [0, \pi]$.

You probably know how to do this. A function can attain the minimum either at one end of the interval or somewhere in the middle. As far as the ends are concerned, the students have already looked a them (their paths correspond to the values $\phi = 0$ and $\phi = \pi$). But how should they look for the minimum inside the interval? Here the universal method is given by differential calculus. Let's take the derivative:

$$S' = \frac{R^2\phi}{\sqrt{H^2 + R^2\phi^2}} - R\sin\frac{\phi}{2}.$$

The points of the interval $(0, \pi)$ where this derivative vanishes, or is not determined (although this is impossible here), are *suspected of being the extremum*. If the function has local maxima or minima in this segment, it can attain these values only in points of this sort. So let's put the derivative equal to zero:

$$\frac{R^2\phi}{\sqrt{H^2 + R^2\phi^2}} - R\sin\frac{\phi}{2} = 0,$$

or, after simplifying,

$$\frac{R\phi}{\sqrt{H^2 + R^2\phi^2}} = \sin\frac{\phi}{2}. \qquad (1)$$

Now all we need to do is find all ϕ satisfying this equation that belong to the interval $(0, \pi)$, calculate the corresponding values of S, and, finally, from all the numbers $S(\phi)$ obtained in this way, as well as $S(0)$, $S(\pi)$, choose the smallest. No sweat!

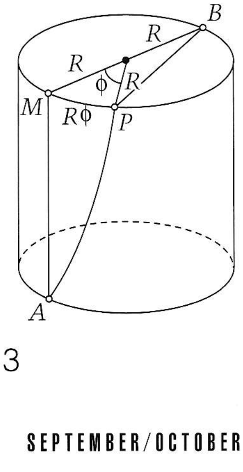


Figure 3

That's easy for *you* to say!

True. The first insurmountable obstacle is to solve equation (1). In fact, *it can't be solved*—that is, it's impossible to express ϕ in terms of H and R by means of elementary functions.

So what shall we do? There's only one way out of this situation: find a "roundabout" method. First, let's note that if a function has a local minimum at a point, then it must decrease to the left of this point and increase to the right. Therefore, its derivative is negative to the left of this point, vanishes at the point, and is positive to the right of it. In other words, the derivative increases in some neighborhood of the local minimum. Thus the second derivative must be positive (or zero, in the extreme case) at the minimum point. And if the second derivative is negative, we can state with certainty that there is no local minimum at the point (it's quite clear there is a maximum).

So, let's find the second derivative:

$$S'' = \frac{R^2H^2}{\left(\sqrt{H^2 + R^2\phi^2}\right)^3} - \frac{R}{2}\cos\frac{\phi}{2}. \qquad (2)$$

We have to find its sign is at the points where the first derivative vanishes—that is, in the points for which equation (1) holds. But how can we do this? Let's try this "trick": we transform the quantities $\sqrt{H^2 + R^2\phi^2}$ and H^2 using equation (1):

$$\sqrt{H^2 + R^2\phi^2} = \frac{R\phi}{\sin\frac{\phi}{2}},$$

$$H^2 = \left(\frac{R\phi}{\sin\frac{\phi}{2}}\right)^2 - R^2\phi^2$$

$$= R^2\phi^2\left(\frac{1}{\sin^2\frac{\phi}{2}} - 1\right)$$

$$= R^2\phi^2\frac{\cos^2\frac{\phi}{2}}{\sin^2\frac{\phi}{2}}.$$

We substitute these formulas in equation (2):

$$S'' = \frac{R^2 \cdot R^2\phi^2 \dfrac{\cos^2 \dfrac{\phi}{2}}{\sin^2 \dfrac{\phi}{2}}}{\left(\dfrac{R\phi}{\sin \dfrac{\phi}{2}}\right)^3} - \frac{R}{2}\cos\frac{\phi}{2}$$

$$= \frac{R}{2\phi}\cos\frac{\phi}{2} \cdot \left(2\cos\frac{\phi}{2}\sin\frac{\phi}{2} - \phi\right)$$

$$= \frac{R}{2\phi}\cos\frac{\phi}{2}(\sin\phi - \phi).$$

If $0 < \phi < \pi$, the factor to the left of the parentheses is positive, and thus the sign of S'' coincides with that of the expression $\sin\phi - \phi$. But, as is well known, $\sin\phi < \phi$ when $0 < \phi < \pi$, and thus $\sin\phi - \phi < 0$ when $\phi \in (0, \pi)$. So if the function S has an extremum in this interval, it can only be a maximum and not a minimum.

The students were right after all!

We see that S attains its minimal value at one of the ends of the segment $[0, \pi]$. This means the students found the correct answer. They were right!

How do you like that? Yet our objections were correct, too. A paradox, wouldn't you say?

No, not really. Just good luck. The students were fortunate in that their incorrect solution gave the correct answer to the problem. Situations like this are not all that rare. You probably can recall something similar that happened to yourself. As a matter of fact, this article was written just to give you a piece of advice (which you may have heard already, but it bears repeating): even when you use the most obvious and the most reliable methods (for instance, the method of planar development, when you're looking for the shortest route), you ought to be very cautious and ready to doubt your reasoning. Otherwise you might get into trouble.

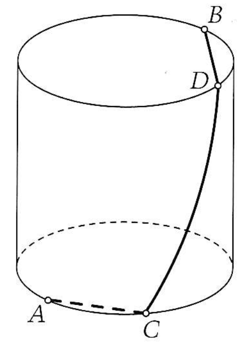


Figure 4

By the way, our solution isn't quite complete either. We didn't take into consideration paths that like the one in figure 4. Think of what to do about them. Then check your answer (in the back of the magazine, as usual).

Physics in the kitchen

Simple experiments with boiling water

by I. I. Mazin

TO PERFORM THE FOLLOWING experiments we need an empty glass, a pan of water, a thermos, a kettle, an electric stove—and what else? Certainly the most important thing: an inquisitive mind, plus the desire to perform some physical "tricks."

Why is water pulled into the glass?

Take an ordinary pan and pour some water into it (to a depth of 2–3 cm). Then lower an empty glass upside down into the pan. Set the pan on the stove, heat the water, and let it boil for about 5 minutes. Turn the stove off. Soon you'll see water being drawn up into the glass, rising higher and higher until it fills most of it.

Now let's explain what we've seen. What is the force that lifts the water in the glass? Clearly it can only be the force of atmospheric pressure (that is, the pressure of the surrounding air). This means that the air pressure inside the glass is less than the atmospheric pressure. By how much? It's not difficult to estimate the pressure difference (ΔP) inside and outside the glass—it's equal to the hydrostatic pressure of the water in the glass at the end of the experiment. Assume the height of the water column is $h \cong 10$ cm, the water density $\rho = 10^3$ kg/m^3, and the acceleration due to gravity $g = 10$ m/s^2. From this we get

$$\Delta P = \rho g h \cong 10^3 \text{ Pa} \cong 0.01 \text{ atm}.$$

So, why did the air pressure inside the glass become less than the atmospheric pressure? The first thought that comes to mind is this: as the water boils, the air in the glass is heated, so it expands and partially leaves the glass. Indeed, looking closely at the pan, we notice air bubbles leaving the glass. When the remaining air cools (after we turn the stove off), it compresses, so the vacated space will be occupied by water. Let's estimate the magnitude of this effect.

Assume the volume of the glass is $V = 200$ cm^3, the initial temperature (before heating) $T_1 = 300$ K, the final temperature $T_2 = 373$ K, and the atmospheric pressure $P = 1$ atm $\cong 10^5$ Pa. From the ideal gas law we obtain the fraction of air that remains in the glass after boiling:

$$PV = \frac{m_1}{M} R T_1,$$

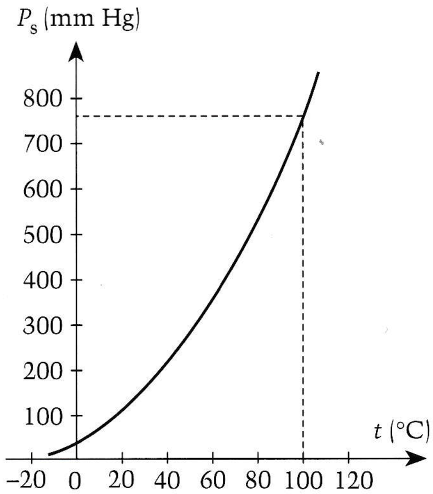
$$PV = \frac{m_2}{M} R T_2,$$

from which we get

$$\frac{m_2}{m_1} = \frac{T_1}{T_2} \cong 0.8.$$

Thus the cooling of the hot air to the initial temperature results in a compression to 80% of the volume of the glass. Thus only 20% is occupied by water. And yet we saw that water filled more than half the glass! At best we can explain only one third of the effect. If we take into account that the water rises in just a few seconds, and that this time is too short for the air to cool to room temperature, we're forced to admit that our explanation is erroneous and that we need to look for another.

Where did we go wrong? It seems we were mistaken in supposing that the glass is filled only with air. We forgot about water vapor. Indeed, during the five minutes of turbulent boiling, water vapor was entering the glass continuously, mixing with the air and trying to push it out. When we turned the stove off, the glass was mostly filled not with air but with water vapor. And not just water vapor—*saturated* vapor. The saturated vapor pressure P_s decreases during cooling, and the drop in pressure is very abrupt (see the figure below). We need to cool the



Art by Pavel Chernusky

water by only 0.3°C to decrease the pressure by 0.01 atm. Clearly such a cooling can take place almost instantaneously.

Our experiment also shows that if the water is boiled long enough, on cooling it will fill almost the entire volume of the glass. There is virtually no limit on the height of the water column—after all, a pressure of 1 atm is created by a water column 10 m high.

The question arises: is five minutes enough time to evaporate the necessary amount of water? Let's try to come up with an answer. The evaporation rate depends on the power output of the stove, the size of the pan, and so on, so let's use the actual numbers obtained in experiments. In our case a layer of water 1 cm deep evaporated from the pan in about 30 minutes. So during five min, $m \cong 3$ g of water will evaporate from a surface equal to the cross-sectional area of the glass—about 20 cm^2. At a temperature T = 373 K and pressure P = 1 atm, this saturated vapor occupies the volume

$$V = \frac{m}{M}\frac{RT}{P} \cong 5 \text{ liters!}$$

Assuming that the vapor mixes homogeneously with the air, we get only (0.2 l/5 l) · 100% = 4% of the volume of the glass is occupied by the air, while the other 96% is occupied by water. This result can actually be observed.

When is it more difficult to pull the cork from a thermos?

For the second experiment we need a thermos, preferably with a narrow cork that plugs the neck tightly but doesn't go into it entirely. First we boil water in a kettle, then pour it into the thermos. A little later we pour it out and plug the thermos tightly with the cork. Try to pull the cork from the thermos a few hours later—you will see how tightly it is set in the neck. You won't have an easy time removing it! (This is why we should use a cork that extends a bit from the neck.) If instead of plugging an empty thermos we had plugged one filled with boiling water, the suction effect would be very small or absent entirely. But what will happen if we fill only a half or a quarter of the thermos with the boiling water? You might think that the suction force would be somewhere in between the two extremes. But that's not the case. The effect will be just a little stronger than with a full thermos. Let's see why.

A characteristic property of a thermos is its low heat transfer. In a good one-liter thermos, water cools only 2–3°C per day. Since the specific heat of water is 4.2 kJ/(kg · K), we can estimate the amount of heat dissipated by the thermos in one day ($\cong$ 10 kJ). An empty thermos has a mass of about 200 g. The specific heat of the thermos's material (glass and metal) is about 0.5 kJ/(kg · K), so a heat transfer of 10 kJ corresponds to a temperature drop of about 100°C. This means that the thermos will cool to room temperature—that is, the temperature drop will be 80°C. According to Charles's law, this temperature drop corresponds to a pressure difference of 0.2 atm. For a cork with a cross section of approximately 5 cm^2, this results in a rather appreciable force of about 10 N.

Since the specific heat of water is almost 10 times that of glass, even 100 g of water in the thermos will reduce the decrease in temperature (and the respective pressure difference) by a factor of four. For a quarter-filled thermos, the pressure difference will be smaller by a factor of seven; in a half-filled thermos, it will be smaller than that for an empty thermos by a factor of 15.

Explain this!

Here's one last experiment. Fill half a thermos with very hot milk, plug it with a cork, and shake it vigorously. You'll see milk bubbles around the cork—air is escaping from the thermos. Can you figure out why? ◘

Bulletin Board

International Mathematical Olympiad

Competing against teams representing a record 82 countries, a team of six US high school students won six medals at the 38th International Mathematical Olympiad (IMO) held in Mar del Plata, Argentina, July 18–31, 1997, and tied for fourth place.

The top 10 teams and their scores (out of a possible 252 points) were China (223), Hungary (219), Iran (217), United States (202), Russia (202), Ukraine (195), Bulgaria (191), Romania (191), Australia (187), and Vietnam (183).

The 1997 IMO US team members were Carl J. Bosley (Topeka, Kansas)—gold medalist, Nathan G. Curtis (Alexandria, Virginia)—gold medalist, Li-Chung Chen (Cupertino, California)—silver medalist, John J. Clyde (New Plymouth, Idaho)—silver medalist, Josh P. Nichols-Barrer (Newton Center, Massachusetts)—silver medalist, and Daniel A. Stronger (New York City)—silver medalist.

Carl Bosley was one of four students (out of 460 participants) who scored a perfect paper.

The Head Coach and Leader of the Team was Titu Andreescu of the Illinois Mathematics and Science Academy. The team was also accompanied by Elgin Johnson of Iowa State University and Walter E. Mientka of the University of Nebraska–Lincoln.

Here is a representative question that was used in this year's IMO:

An $n \times n$ matrix (square array) whose entries come from the set $S = \{1, 2, \ldots, 2n - 1\}$ is called a silver matrix if, for each $i = 1, \ldots, n$, the ith row and the ith column together contain all elements of S. Show that (a) there is no silver matrix for $n = 1997$; (b) silver matrices exist for infinitely many values of n.

The American Mathematics Competitions (AMC) is a program of the Mathematical Association of America, and the USA Mathematical Olympiad is an AMC activity sponsored by nine national mathematical sciences organizations. Financial and progam support is provided by the Army Research Office, the Office of Naval Research, Microsoft Corporation, the Matilda R. Wilson Fund, and the University of Nebraska–Lincoln.

International Physics Olympiad

A report on the XXVIII International Physics Olympiad will appear in the November/December issue.

Not a plane old CyberTeaser

The September/October CyberTeaser (brainteaser B212 in this issue) was a cinch—once you realized you weren't going to be able to solve it by pushing matchsticks around the tabletop. We were impressed by the ASCII drawings sent in by some of our contestants (the CyberJudge is all thumbs in that department) and by the JPEGs and bitmaps as well. But of course, all you needed was the right words . . .

Here are the first ten persons who submitted a correct answer electronically:

Theo Koupelis (Wausau, Wisconsin)
Matthew Wong (Edmonton, Alberta)
Brian S. Mansfield (Loveland, Ohio)
Leo Borovskiy (Brooklyn, New York)
Jim Grady (Branchburg, New Jersey)
Badri Ramamurthi (Albuquerque, New Mexico)
Chantelle March (Morphett Vale, South Australia)
Jim Paris (Doylestown, Pennsylvania)
Lee Ai Ling (Darul Ridzuan, Malaysia)
Oleg Shpyrko (Cambridge, Massachusetts)

Each will receive a *Quantum* button and a copy of the September/October issue. Everyone who submitted a correct answer in the time allowed was eligible to win a copy of *Quantum Quandaries*, our collection of the first 100 *Quantum* brainteasers.

Care to go toe to toe with the latest CyberTeaser? Then head to www.nsta.org/quantum and click on the Contest button.

What's happening?

Summer study ... competitions ... new books ... ongoing activities ... clubs and associations ... free samples ... contests ... whatever it is, if you think it's of interest to *Quantum* readers, let us know about it!

What's on your mind?

Write to us! We want to know what you think of *Quantum*. What do you like the most? What would you like to see more of? And, yes—what *don't* you like about *Quantum*? We want to make it even better, but we need your help.

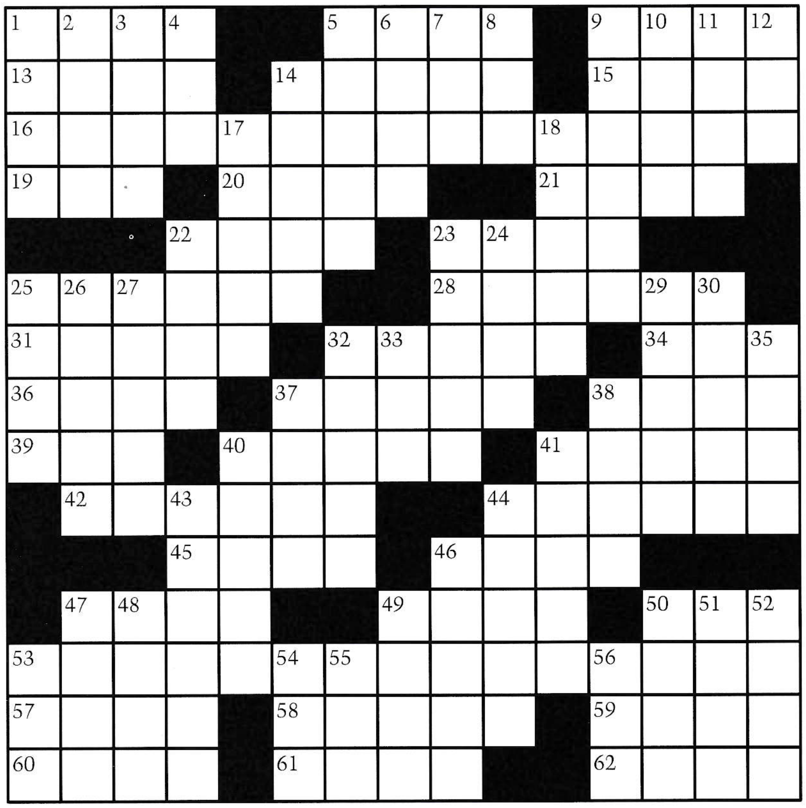
What's our address?

Quantum
National Science Teachers Association
1840 Wilson Boulevard
Arlington VA 22201-3000

Be a factor in the QUANTUM equation!

Criss × cross science

by David R. Martin



Across

1 German physicist Ernst ___ (1840–1905)
5 Fusible ceramic mixture
9 60,091 (in base 16)
13 Move suddenly
14 Instant
15 Great ___ (dog)
16 Like a battery
19 Sum
20 Mongoloid or Caucasian, e.g.
21 Spore sacs
22 Test tube baby pioneer John ___
23 Luminous ring
25 Like the ocean
28 Organic compounds
31 Muslim doctors
32 Excessive
34 ___ bang
36 BBQ favorite
37 Regime
38 ___-Civita symbol
39 Abscisic acid: abbr.
40 Elongated fruits
41 Laser infrared radar
42 Polysaccharide
44 Trig. function
45 Absorbed dose units
46 Wheel hub
47 Birth-control advocate ___ Guttmacher
49 Enclosure
50 ___ wind (of western China)
53 Selective diffusion process
57 Jai ___
58 Austrian composer ___ Berg
59 Anthropologist ___ Hrdlicka
60 Rotate
61 Nevada lake
62 Unit of heredity

Down

1 44,522 (in base 16)
2 Hairless
3 Procreated
4 And so on: abbr.
5 Dress
6 Food grain
7 Fish dermatitis
8 "to a ___." (perfect)
9 Inventor Thomas ___
10 43,724 (in base 16)
11 ___ b'rith
12 Ten decibels
14 Sum of diagonal matrix elements
17 Sodium sesquicarbonate
18 ___ acid (green apple juice ingredient)
22 Edges or borders
23 Suspends
24 60 coulombs: abbr.
25 Koran chapter
26 Pretext
27 Rachel and Leah's father
29 966,362 (in base 16)
30 Jewish month
32 Averages
33 Atmosphere
35 Fastened with a belt
37 Geophysicist Harry Fielding ___ (1859–1944)
38 Mallophaga
40 Type of kingdom
41 Surveyor's instrument
43 $C_4H_4N_2O_2$
44 Astronomer Carl ___
46 Nymph
47 More stable isomer: pref.
48 Loyal, in Edinburgh
49 52,666 (in base 16)
50 Seagirt land
51 Tres ___
52 Exist in France
53 Sense organ
54 Computer memory: abbr.
55 Novelist ___ Edvart Rolvaag (1876–1931)
56 ___ laser

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE JULY/AUGUST PUZZLE

R	E	C	K	■	X	Y	L	A	N	■	T	A	A	L
U	S	E	E	■	R	E	A	D	Y	■	E	E	A	A
P	A	L	P	■	A	A	B	A	A	■	T	B	A	R
P	U	L	L	E	Y	S	■	D	S	E	R	I	E	S
■	■	■	E	S	T	■	T	E	S	L	Å	■	■	■
A	P	E	R	T	U	R	E	■	A	L	P	A	C	A
O	R	T	■	E	B	O	N	Y	■	S	O	D	A	S
R	O	H	E	■	E	O	S	I	N	■	D	O	S	T
T	W	Y	L	A	■	T	O	P	I	C	■	N	E	E
A	L	L	E	L	E	■	R	E	C	E	P	T	O	R
■	■	■	C	O	R	D	S	■	H	E	R	■	■	■
P	E	N	T	E	N	E	■	A	R	S	E	N	I	C
A	G	A	R	■	E	L	A	E	O	■	L	E	D	A
L	E	N	O	■	S	A	L	A	M	■	O	V	I	D
I	R	O	N	■	T	Y	P	E	E	■	G	E	N	E

ANSWERS, HINTS & SOLUTIONS

Math

M211

The set defined by the equation $|y - 2x| = x$ consists of two rays: $y = x$, $x \geq 0$, and $y = 3x$, $x \geq 0$. It's not hard to see that the equation $|3x - 2y| = y$ defines the same rays. There are no other possibilities.

M212

Triangle CPK is congruent to triangle CBK, because CK is a common side (fig. 1), $\angle PCK = \angle BCK$, and $\angle KPC = \angle KMA = \angle CBK$. The last two equalities follow from the properties of inscribed angles. Thus $AP = |CP - CA| = |a - b|$.

M213

The system is linear with respect to parameters a, b and c. Let's use it to express a, b, and c via x, y, and z.

Eliminate denominators in the second and the third equations and replace the expression given in the first equation with c. After simplifying, we obtain

$$\begin{cases} -a\left(1+x^2\right)yz + b(xz+y)x \\ \qquad = x^2z^2y - x^3y^2z, \\ a(y - xz)z - b\left(1+z^2\right)xy \\ \qquad = -x^2z^2y - xy^2z^3. \end{cases}$$

Get rid of b (multiplying the equations by $(1 + z^2)y$ and $(xz + y)$, respectively, and adding the first equation to the second). We obtain

$$(1 + x^2)(1 + z^2)y^2z + (y - xz)(y + xz)z)$$
$$= xyz((xz - x^2y)(1 + z^2)y - (xz + yz^2)(xz + y)),$$

or

$$a(-x^2y^2z^3 - x^2y^2z - y^2z^3 - y^2z + y^2z - x^2z^3) = xyz(xyz + xyz^3 - x^2y^2 - x^2y^2z^2 - x^2z^2 - xyz - xyz^3 - y^2z^2)$$

—that is,

$$az(x^2y^2z^2 + x^2y^2 + y^2z^2 + x^2z^2)$$
$$= xyz(x^2y^2z^2 + x^2y^2 + y^2z^2 + x^2z^2).$$

But it's evident that $x, y, z \neq 0$. Thus $a = xy$. Further, we find $b = yz$, $c = xz$. We thus obtain the system

$$\begin{cases} a = xy, \\ b = yz, \\ c = xz. \end{cases}$$

Multiplying these three equations term by term, we find $abc = x^2y^2z^2$. Thus $abc > 0$, and $xyz = \pm\sqrt{abc}$. Therefore, the answer is

$$\left(\frac{\pm\sqrt{abc}}{b}, \frac{\pm\sqrt{abc}}{a}, \frac{\pm\sqrt{abc}}{c}\right).$$

M214

Consider a triangle ABC that can be folded into the surface of a unit regular tetrahedron (without overlaps), so that the vertices of the tetrahedron correspond to the vertices of the triangle and to the midpoints of its sides. Draw the median AD in the triangle and continue it after point D to a distance equal to itself. We obtain point E. Then we can fold the triangle ABE into the surface of the same tetrahedron so that the vertices of this tetrahedron correspond to the vertices of the triangle and to the midpoints of its sides. (This fact is evident enough. Triangle ACD is, in a certain sense, substituted for triangle BED, which takes the place of ACD on the surface of the tetrahedron.)

Figure 2 shows how we can obtain, using the transformation above, the triangles described in parts (a) and (b) from an equilateral triangle with side 2. This triangle is a development of a unit regular tetrahedron, so that the vertices of the tetrahedron correspond to the vertices and midpoints of the sides of the triangle. Thus we conclude that it is possible to fold these triangles into the surface of a unit regular tetrahedron.

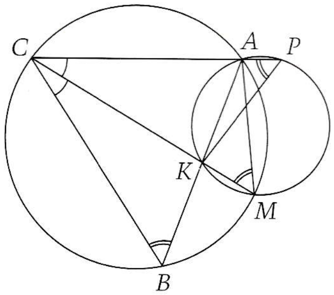


Figure 1

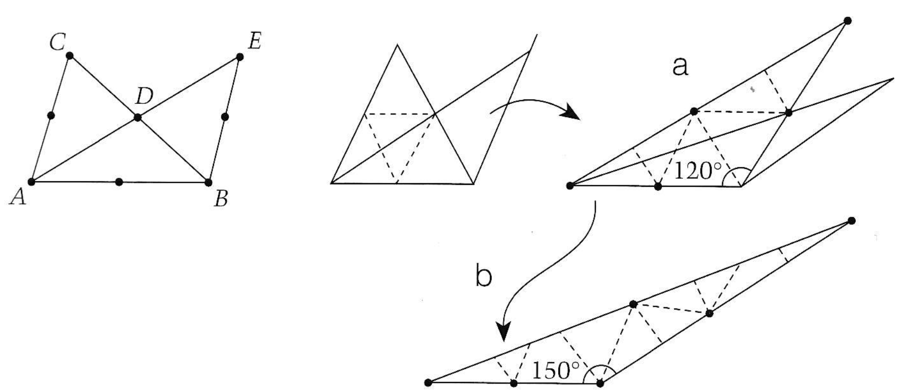


Figure 2

M215

Note that $b(2n + 1) = b(n)$. (The first digit in any such representation of $2n + 1$ is 1. Dropping it and dividing all the rest by 2, we obtain a representation of n. Thus we get a one-to-one correspondence between the representations of n and $2n + 1$.) We also note that $b(2n) = b(n) + b(n - 1)$ (the first term corresponds to the case $a_0 = 0$, and the second to the case $a_0 = 2$). Now we compute

$$b(1997) = b(998) = b(499) + b(498) = 2b(249) + b(248) = 3b(124) + b(123) = 3(b(62) + b(61)) + b(61) = 3b(62) + 4b(61) = 3b(31) + 7b(30) = 10b(15) + 7b(14) = 17b(7) + 7b(6) = 24b(3) + 7b(2) = 31b(1) + 7b(0) = 31 \cdot 1 + 7 \cdot 1 = 38.$$

Physics

P211

At any moment while the athletes are running the cord is stretched uniformly, so the ratio of the distances from the point C to the ends of the cord will not vary with time. The figure in the problem shows that this ratio is initially equal to

$$|AC| : |CB| = 1 : 4.$$

Clearly the displacement Δx of the knot to the east is determined by the displacement ΔS_x of runner A—at all times it equals 4/5 of that displacement:

$$\Delta x = \frac{4}{5}\Delta S_x = \frac{4}{5}v_0 t.$$

Using the dimensions given in the figure, we see that point D is $\Delta x = 4$ m to the east. Therefore, the knot passes through point D when

$$t = \frac{5}{4}\frac{\Delta x}{v_0} = 5\text{ s}.$$

The displacement Δy of the knot to the south is determined by the shift ΔS_y of runner B, so at any moment $\Delta y = \Delta S/5$. At time $t = 5$ s the knot is shifted from the initial position to the south by $\Delta y = 2$ m. Thus runner B, moving with an acceleration a, runs the distance $\Delta S_y = 5\Delta y = 10$ m in time $t = 5$ s—that is,

$$\frac{1}{2}at^2 = \Delta S_y,$$

which reduces to

$$a = 2\frac{\Delta S_y}{t^2} = \frac{20}{25}\text{ m/s}^2 = 0.8\text{ m/s}^2.$$

P212

We'll denote the temperature and pressure of the air at a distance from the model as T_1 ($T_1 = T = 300$ K) and P_1, and the corresponding values near point A as T_2 and P_2. For a stationary flow we can consider any portion of the gas—how it moves and what happens to it. For definiteness, let's take one mole of air (the molecular mass of air $M = 29$ g/mole) and look at a "tube" that it enters from far off and from which it emerges near the model. To avoid purely formal difficulties associated with the complete stoppage of the air near our point, we'll consider this speed small compared to the initial speed (but not exactly equal to zero!).

The outside air, "pushing" our portion of gas into the "tube" at the entrance, performs work

$$W_1 = P_1V_1,$$

where V_1 is the volume of one mole of air at temperature T_1. Leaving the tube, the gas performs work, "pushing away" the surrounding air—that is, negative work is performed on the gas:

$$W_2 = -P_2V_2,$$

where V_2 is the volume of a mole of gas at temperature T_2.

Let's assume that the gas in the "tube" does not exchange heat with the surrounding air. (Strictly speaking, this isn't so, but we can't reasonably estimate this heat exchange. Therefore, we'll obtain the upper limit for the temperature effect we're investigating.) The change in the internal energy of our portion of gas is determined by the work done by the outside forces and the change in the kinetic energy of this portion of gas as a whole:

$$\begin{aligned} W_1 + W_2 + \frac{Mv^2}{2} &= \Delta U \\ &= C_V(T_2 - T_1) \\ &= \frac{5}{2}R(T_2 - T_1) \end{aligned}$$

(since air is a diatomic gas, its molar heat capacity at constant volume is $C_V = {}^5/_2R$).

Using the equation of state for one mole of ideal gas, $PV = RT$, we find

$$W_1 = P_1V_1 = RT_1,$$
$$W_2 = -P_2V_2 = -RT_2.$$

Now we can write

$$RT_1 - RT_2 + \frac{Mv^2}{2} = \frac{5}{2}R(T_2 - T_1),$$

from which we get

$$T_2 = T_1 + \frac{Mv^2}{7R} \cong 345\text{ K}.$$

P213

For temperature measurements the thermometer must be heated from room temperature to that of the human body—that is, by about 15–17°C. The mercury in the thermometer can be shaken down when its temperature drops by 3–4°C. Since the thermometer's scale begins at 34°C, a temperature decrease of a few degrees produces empty space above the crimp. We must take into account that when objects are heated and cooled, the rates of their temperature changes are proportional to the temperature difference between the object and its surroundings, so the dependence of a thermometer's temperature on time looks like the curve in figure 3. Therefore, the time necessary to cool a thermometer to the temperature at which the mercury can be

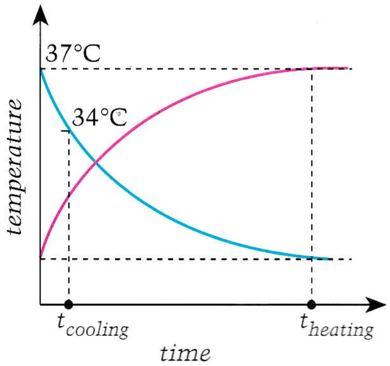


Figure 3

shaken down is far less than the time needed to measure the body temperature.

P214

Let's give the system of splinters an electric charge q ($q > 0$ for definiteness) and begin to reconstruct the sphere from the fragments connected by the wires (whose capacitance we ignore). Clearly the surface charge density at any part of the recomposed sphere will be positive. Thus the electrostatic forces cause the fragments to repel each other the entire time they are near one another. So we need to perform work $W > 0$ to restore the sphere.

The charge of the restored sphere is q and its electrostatic energy is $E_s = q^2/2C_s$, where C_s is the sphere's capacitance. The energy of the system of fragments was $E_f = q^2/2C_f$, where C_f is its capacitance. Clearly $W = E_s - E_f$, and since $W > 0$, then $E_s > E_f$—that is,

$$\frac{q^2}{2C_s} > \frac{q^2}{2C_f}.$$

Therefore, $C_s < C_f$—that is, the electric capacitance of the original sphere is less than the total capacitance of its fragments connected by wires.

P215

The spot of light that can be seen in the large mirror is a reflection of the spot of light formed on a "screen" in front of the large mirror. The role of the screen may be played by the observer's body, a wall reflected in the mirror, and so on. However, the screen has a "hole" in it—the small mirror itself. If the sunlight reflected first from the small mirror, and then from the large one, strikes the small mirror again, there will be no spot of light. After the first reflection from the small mirror, the beam passes perpendicular to the plane of the large mirror, and thus it hits the image of the small mirror in the large one. Therefore, if the observer directs her spot of light at the image of the small mirror, she will no longer see the spot of light.

Brainteasers

B211

Arrange the matches in a cube with an edge length of one match.

B212

One possible answer is to make the six points the vertices of two equilateral triangles with unit side length, lying at a unit distance from each other (fig. 4).

B213

Let x denote the number of people who are both mathematicians and philosophers. Then the number of mathematicians is $7x$, and the number of philosophers is $9x$. Thus philosophers are more numerous. (One might wonder whether x is zero. But if this question has occurred to you, then you are both a mathematician and a philosopher, and thus x is different from zero.)

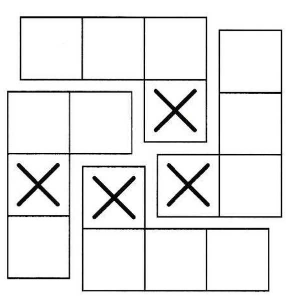


Figure 4

B214

Our task is to find the smallest natural n such that one can find a whole number between $96/35 \cdot n$ and $97/36 \cdot n$. It's not hard to check that the desired number is 7, and the fraction 19/7 ($96/35 \cdot 7 \geq 19 \geq 97/36 \cdot 7$). We can verify by direct computation that numbers smaller than 7 don't work.

B215

See figure 5.

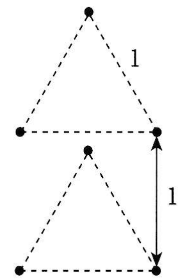


Figure 5

LTG revisited

1. The value of \$100 is given in the table below. Continuing the process for a few more decades is likely to convince you that your balance will never grow past \$200.

Time	Balance	Interest – fee
0–10 years	100	100 – 50 = 50
10–20 years	150	150 – 112.50 = 37.50
20–30 years	187.50	187.50 – 175.78 = 11.72
30 years	199.22	

2. The value of \$100 is given in the table below. This form of sloppy bookkeeping leads to balances that oscillate about \$200 and correspond to the phenomenon of "chaos."

Time	Balance	Interest – fee
0–30 years	100	300 – 150 = 150
30–60 years	250	750 – 937.50 = –187.50
60–90 years	62.50	187.50 – 58.59 =128.91
90–120 years	191.41	574.22 – 549.55 = 24.67

3. In figure 14 (in the article), we need only draw a connector from the rectangular reservoir labeled **Popula-**

tion P to the circular converter labeled **c**. Of course we must also assign a value to **L** and program the rule "**c = L/P(t)**" into the system. This is done by means of a dialog box that is opened by double-clicking the icon **c**.

4. One possibility is $c = L(2 - P)/P$. Another is $c = L(2 - P)/\sqrt{P}$. Which do you think is more realistic? Why?

Virus

Why will there be a stabilization of the number of INFECTED in the case of the disease spreading within one population? What will that number of INFECTED be in the long run? The model shows initially a severe outbreak of the disease because the initial stocks of NONIMMUNE and INFECTED people are large. Thus, the product of

RATE OF CONTACT * NONIMMUNE * INFECTED

is large. As the infected move on to either die or survive, the stocks become smaller. In the long-run, the only NONIMMUNE ones in the population are the NONIMMUNE IMMIGRANTS, to whom the disease will be passed on. Thus, the number of INFECTED will be, in the long run, those seven NONIMMUNE IMMIGRANTS.

Why does the disease entirely disappear in the two-population model when the virus is passed on by direct contact? With a lower contact rate, the stock of NONIMMUNE H can temporarily build up again after the first severe outbreak of the disease, and the stock of INFECTED H becomes larger for each subsequent outbreak. The more individuals that are infected, the more are being removed in the following period from the system, and INFECTED H reaches zero. Similarly, the INFECTED M rapidly goes to zero, and as a result

H RECEIVE VIRUS = RATE OF CONTACT H1 * NONIMMUNE H * INFECTED H + RATE OF CONTACT H2 * NONIMMUNE H * INFECTED M = 0.

One of the ways in which the disease can reenter the population is by reappearing in a mutated form. Can you model that case?

For a powerful description of the dynamics of the Ebola virus, see *The Hot Zone* by R. Preston (New York: Anchor Books, 1995).

World in a bubble

1. For the Biosphere 2 rainforest, the residence time of carbon in a reservoir (T) = Reservoir size (gC)/Rate of inflow into or outflow from reservoir (gC/unit time).

- For the atmosphere: T = (50,000 gC)/(4,000 gC/hr) = 12.5 hours
- Plants: T = (1,100,000 gC)/(4,000 gC/hr) = 275 hours = 11.5 days
- Soils: T = (100,700,000 gC)/(2,000 gC/hr) = 50,350 hours = 2,100 days = 5.75 years

2. Global carbon cycle:

- Atmosphere: T = (615 Gt)/(124 Gt/yr + 60 Gt/yr) = 3.3 years
- Plants: T = (731 Gt)/(124 Gt/yr) = 5.9 years
- Soils: T = (1,238 Gt)/(62 Gt/yr) = 20 years
- Ocean: T = (36,866 Gt)/(60 Gt/yr) = 614.4 years

3. • Step 1. Find the amount of Gt of air in the Earth's atmosphere:

$(0.028/100\ CO_2) \cdot x$ Gt air = 615 Gt CO_2 in air

$x = 2.2 \cdot 10^6$ Gt air

• Step 2. Determine what a 2 ppm (or 0.0002%) increase in CO_2 is in Gt:

$(0.0002/100\ CO_2$ each year$) \cdot 2.2 \cdot 10^6$ Gt air = x Gt added each year

$x = 4.4$ Gt CO_2 added each year

Ant

The length of the path *ACDB* equals that of the path *AC′DB*—see figure 6. And the latter path is clearly longer than *AC′B*. So our answer is right.

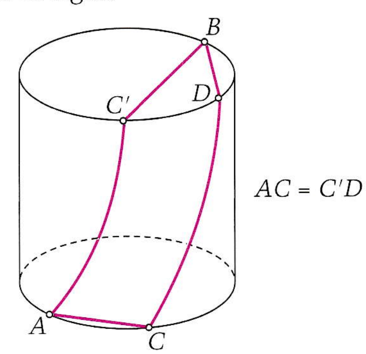


Figure 6

Talk back to us

drop us a line at *Quantum*,
1840 Wilson Blvd.,
Arlington VA 22201

zip us an electron or two at
quantum@nsta.org

Bad milk

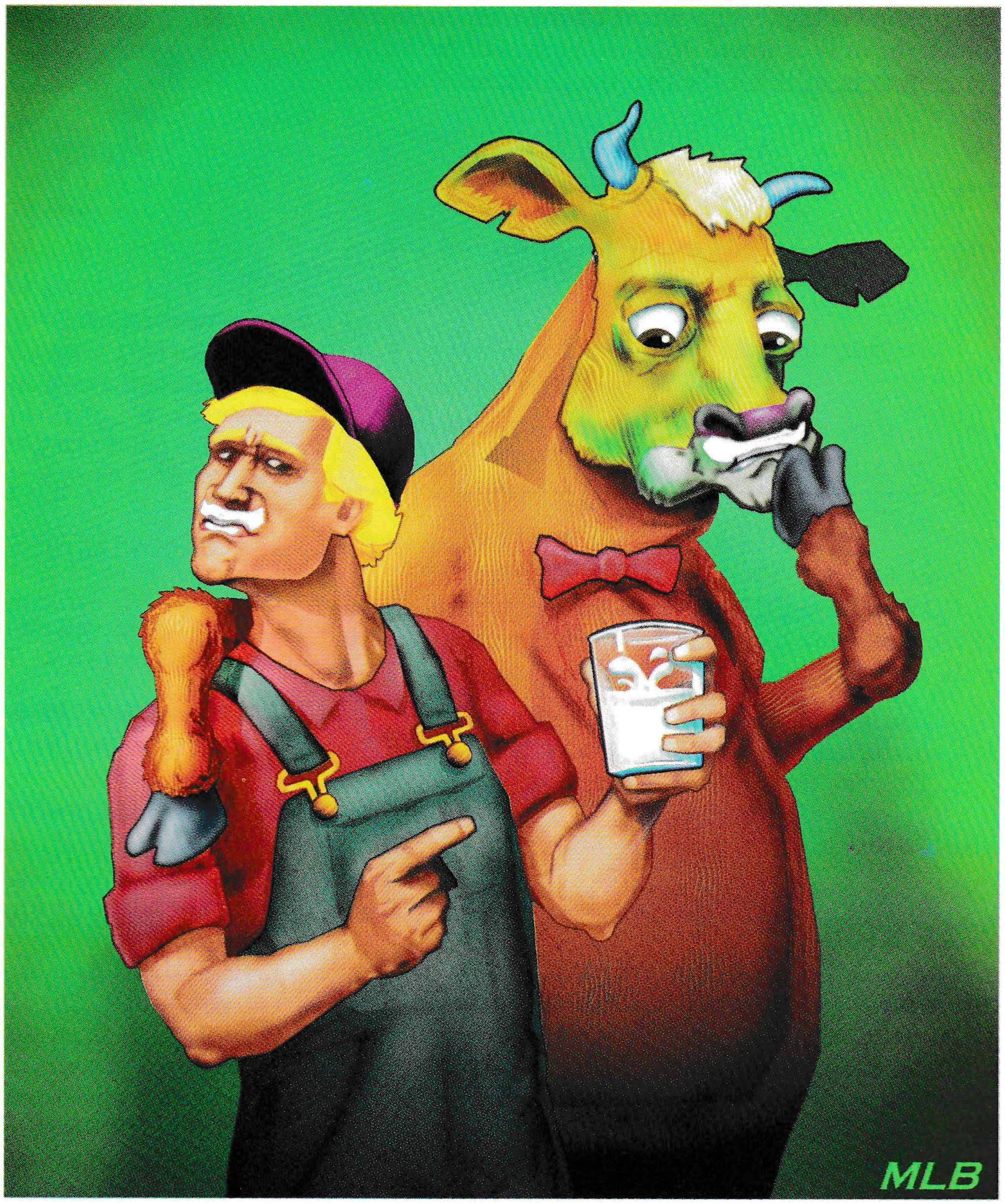
A dynamic system gone sour

by Dr. Mu

WELCOME BACK to Cowculations, the column devoted to problems best solved with a computer algorithm.

The average cow in Farmer Paul's herd produces 10 gallons of milk a day. The raw milk goes directly from the cow through the milking machine into a cooling tank, where it is held until it is picked up by a refrigerated milk truck every day, 365 days a year. Thus, within a day or two, milk that my bovine friends have lovingly produced is on your local supermarket shelf ready for your enjoyment.

Refrigeration is the key to preserving milk. Without it, the shelf life of our sweet natural product would be very short. Bacteria, which enters the milk from many sources—none of which I want to discuss in public—soon starts to multiply and eventually transforms our sweet nectar into a sour mess. The growth of bacteria in milk has been identified by Farmer Paul as obeying a Discrete Dynamical Sys-



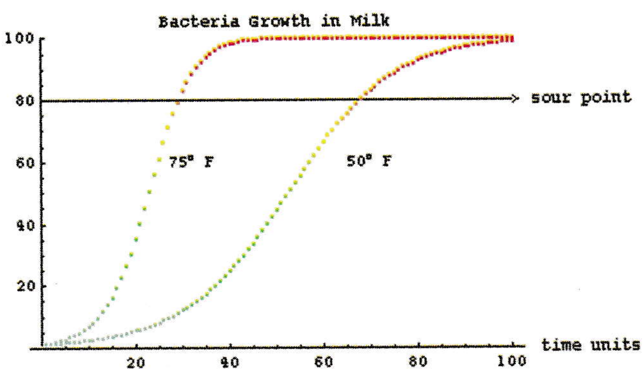
Art by Mark Brenneman

tem (DDS) similar to the models used to predict the population growth of humans. Think of bacteria as little people.

Farmer Paul's DDS Model for bacteria growth in milk states that if we take a series of bacteria measurements in milk at equally spaced times, then the bacteria count changes according to the following Logistic Growth Model (note: if *now* is the present time period, *now* – 1 is the previous time period, with 1 representing a fixed unit of time):

```
Bacteria[0]=1;
Bacteria[now]=Bacteria[now-1] +
(Temperature - 32)
------------------ Bacteria[now-1](1-
       200
Bacteria[now - 1]
-----------------)
       200
```

The length of time it takes the bacteria to reach a reading of 80 is the length of time it takes to sour. This time period is highly dependent on the temperature at which the milk is stored. Here is a picture of the bacteria growth in milk for the two temperatures 50°F and 75°F:



Now for your "Challenge Outta Wisconsin."

COW 6. Given Farmer Paul's model for the growth of bacteria in milk, cowculate the temperature at which milk will sour twice as fast as it does at 50°F.

The milk has gone to the cooling tower.
Better use some computer power.
To crack this COW,
You must know how
To find the time when the milk turns sour.

—Dr. Mu

Solution to COW 5

Last time I proposed the problem of how to fairly award prize money to two baseball teams that end the series before one team wins 50 games.

COW 5. Write a program that will cowculate the winnings of each team based on BABE's rules. You are to assume that if the game score is currently at Holsteins: H, and Jerseys: J, the probability that the next game will be won by the Holsteins is $H/(H + J)$, while the probability it is won by the Jerseys is $J/(H + J)$. Also, if P is the probability that the Jerseys will win fifty games first, then they should be awarded $P \cdot 1{,}000$ of the prize money. Report your answer for the series that ended Holsteins: 35 and Jerseys: 41.

Let **ProbJWins[H,J]** represent the probability that the Jerseys will win 50 games before the Holsteins, given the current games won is H for Holsteins and J for Jerseys. Clearly, **ProbJWins[H,50] = 1** for $H < 50$ because the Jerseys have won 50 games. Also, **ProbJWins[50,J]=0** for $J < 50$ because the Holsteins have won the series. We can work backward from the boundary conditions with the following recursive relationship:

$$ProbJWins[H,J] = \frac{H}{H+J} ProbJWins[H+1,J] + \frac{J}{H+J} ProbJWins[H,J+1].$$

This simply says that the only way to win from the current score of H to J is for the Holsteins to win the next game (with probability $H/(H + J)$) and then to win from a score of $H + 1$ to J, or to have the Jerseys win the next game (with probability $J/(H + J)$) and then win from a score of H to $J + 1$. One of these two results must occur and they are mutually exclusive.

These conditions are specified in Mathematica® as follows:

```
Clear[ProbJWins]
ProbJWins[H_, 50]:=1/;H<50
ProbJWins[50, J_]:=0/;J<50
ProbJWins[H_,J_]:=

                          H
ProbJWins[H,J]=  ------- ProbJWins[H+1,J]
                        H + J

    J
+ ------- ProbJWins[H,J+1]
  H + J
```

The winning amount for the Jerseys, given a series ending at $H = 35$ and $J = 41$, is

```
1000.ProbJWins[35,41]

924.097
```

Not bad, and clearly better for the Jerseys than it would have been if other settlement methods were used.

Note that in this cowculation, past performance is an indicator of future performance, because the $H/(H + J)$-factor—the chance H wins the next game—is more favorable to the Holsteins if they are ahead in the number of wins to date. The classical version of this problem, called the *problem of points*, uses a fixed probability p for the chance of a win at each game. In this case, past performance is not assumed to be an indicator of future performance. If we assume the Holsteins and the Jerseys are an even match in every game, then $p = 0.5$ for each game and we apply the classical model. We get the following set of recursive relationships:

```
Clear[ProbJWins]
ProbJWins[H_,50]:=1/;H<50
ProbJWins[50,J_]:=0/;J<50
ProbJWins[H_,J_]:=ProbJWins[H,J]=.5
ProbJWins[H+1,J]+.5ProbJWins[H,J+1]
```

The winning purse for the Jerseys if the classical model is used, given a series ending at $H = 35$ and $J = 41$, is

```
1000.ProbJWins[35,41]
```

```
894.98
```

Not as big a difference as might be expected—only about $30 less. This classical form was first posed to the French mathematician Pascal in 1654 as a gambling problem: how to split the pot before a game has ended given the present state of the game. A closed form mathematical solution exists for the classical case, but not for COW 5. This one requires a computer algorithm.

A simulation solution

Another way to approach this problem is to actually simulate playing out the series by using random numbers to decide who wins each game. According to BABE's rules, if the present game score is Holsteins H and Jerseys J, then by picking a random number between 0 and 1 (**Random[]**), the Jerseys win the next game if **Random[]<J/(H+J)**; otherwise the Holsteins win. Once a team has reached 50 wins, the series if over. So, we start with the score Holsteins: 35, Jerseys: 41, and record who reaches 50 first. We repeat this 1,000 times and compute the percentage of times that the Jerseys reached 50 first.

Morton Goldberg submitted a compiled solution in Mathematica. Presented below is a variation of his solution, without any compile.

Begin by defining the function **PlayBall**, which takes the current gameboard {Hostein wins, Jersey wins, Goal}, plays one game, and updates the gameboard. Once either side has reached the goal, the gameboard remains the same:

```
PlayBall[s_]:=Module[{h=s[[1]],j=s[[2]]},
    If[Random[]<j/(h+j),{h,j+1,s[[3]]},
  {h+1,j,s[[3]]}]];
PlayBall[s_]:=/;Max[s[[1]],s[[2]]]==s[[3]]
```

A series is over once a team has reached the goal (50 wins) and **PlayBall** remains fixed. This is done in Mathematica with the **FixedPoint** function. Now we need to keep a tally of which side reached the goal of 50 games first. We define the **Tally** function: if the Holsteins reach the goal first, increment by one the number of Holstein series victories; otherwise increment the wins for the Jerseys:

```
Tally[s_]:=If[s[[1]]==s[[3]],Holsteins++,Jerseys++];
```

Now all that remains is to set the variables to zero and define the standings:

```
{Holsteins=0,Jerseys=0};
standing={35,41,50};
```

We repeat the series 1,000 times and print the outcome:

```
Do[Tally[FixedPoint[PlayBall,standing]],{1000}];
{Holsteins,Jerseys,N[Jerseys/
(Jerseys+Holsteins),3]}
```

```
{69,931,0.931}
```

Morton Goldberg averaged the output from 10 independent simulations, resulting in $P = 0.925 \pm 0.003$, giving the Jerseys a prize of $925. This is very close to the analytical solution of $924.10.

And finally . . .

Please end your cowculation for COW 6 to drmu@cs.uwp.edu. To view all previous COW ruminations, take a peek at http://usaco.uwp.edu/cowculations.

Read more about it on the World Wide Web!

Here is a list of World Wide Web sites mentioned in various LTG articles in this issue:

- Biosphere 2: www.bio2.org
- High Performance Systems, Inc. (STELLA): www.hps-inc.com
- Ventana Systems, Inc. (Vensim): www.vensim.com
- Matthias Ruth: web.bu.edu/CEES/readmoreMR.html

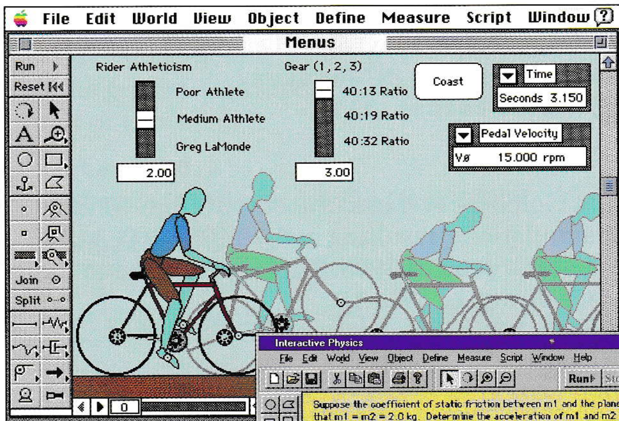
The Club of Rome maintains a Web site (www.ClubOfRome.org) where you can find a history of that organization, a "short version" of *The Limits to Growth*, and many useful links.

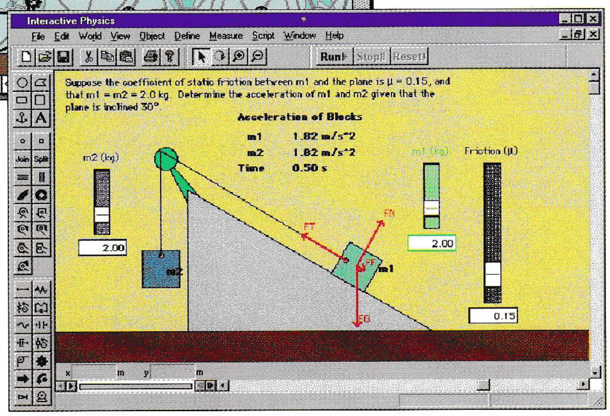
An FTP site at the University of Illinois (ftp.ncsa.uiuc.edu/GlobalModels/SoftWare) has a Mac Hypercard version of the World3 model that includes a set of classroom materials (BeyondTheLimits.Mac.SEA.hqx) and other items of interest.

You can also use a Web search engine to find many other sites, entering such search terms as "system dynamics," "sustainability," "World3," and the like.

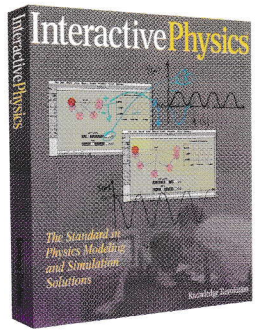
Revolutionize Your Curriculum!

Our Integrated Mathematics and Science Solutions Redefine Your Educational Environment





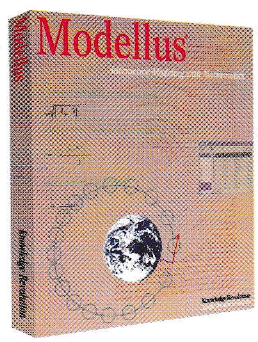
Create simulations by sketching out bodies and attachments with a simple point-and-click user interface. Alter physical properties like friction and mass during the simulation to gain valuable insight into the system.



Interactive Physics™ makes it easy to integrate modeling and simulation into your physics curriculum by offering a complete desktop motion lab. Create models by drawing onscreen with a powerful and easy-to-use graphic interface. Add objects like springs, dampers, ropes and joints. Measure physical quantities like velocity, acceleration, momentum and energy. Interact with your model in real time by changing properties as the simulation runs. Widely adopted by many textbook publishers and with more than half a dozen awards and thousands of educational users, *Interactive Physics* is the standard in physics modeling and simulation solutions.

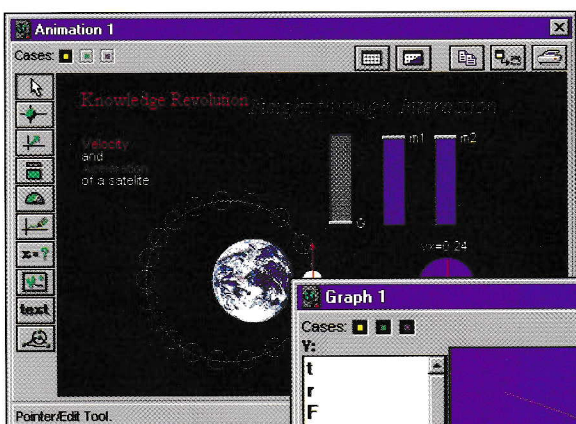
Interactive Physics: Single User $249
Lab Pack pricing . . . . .Call

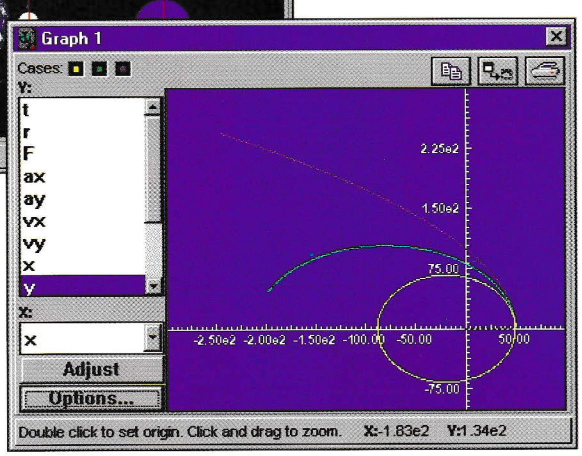
NEW



Modellus™ gives students the power to interact with models in real time to gain powerful insights into the underlying mathematics. Its equation-based approach to modeling offers an educational compliment to *Interactive Physics.* Typical math programs act as calculators, treating the model as a static entity. But *Modellus* lets students manipulate mathematical variables during a simulation to study the effects of their changes on animations, graphs and tables. By supporting multiple cases and multiple data representations, *Modellus* offers students valuable feedback during the problem solving process.

Modellus: Single User $199
Lab Pack pricing . . . . .Call





With Modellus, simulations are created from mathematical equations. Explore scientific models by interacting with algebraic, differential, and iterative equations in the context of a physical system.

Knowledge Revolution

Insight through Interaction

66 Bovet Road, Suite 200
San Mateo, California 94402
Fax 650 574 7541
Telephone 650 574 7777
Toll-Free 800 766 6615

Call Now for Demo CD

www.krev.com

Circle No. 1 on Reader Service Card